International Journal of Interactive Multimedia and Artificial Intelligence

September 2020, Vol. VI, Number 3
ISSN: 1989-1660

unir LA UNIVERSIDAD EN INTERNET

This picture is in honor of Jefferson Cardenas Guerra, author and winner of the call #yomequedoencasadiseñando - #Istayathomedesigning - launched by the Virtual Museum of the Engineering and Technology School of UNIR. 2020

Special Issue on Artificial Intelligence and Blockchain

#YoMeQuedoEnCasa

INTERNATIONAL JOURNAL OF INTERACTIVE MULTIMEDIA AND ARTIFICIAL INTELLIGENCE
ISSN: 1989-1660 –VOL. 6, NUMBER 3

Editor's Note

Towards Blockchain Intelligence

In 2008, Satoshi Nakamoto introduced the Bitcoin [1], the first cryptocurrency that had a significant impact in the financial scene. This currency relied on some technological innovations, including a cryptographic mining system and a distributed database with interesting properties, deemed "the Blockchain".

At its core, Blockchain is a database able to store a list of entries, implementing some cryptographic mechanisms that prevents stakeholders from being able to tamper its contents. This is achieved by storing data in blocks, which are summarized using a hash function and linked to each other (therefore the term "Blockchain"). Additionally, Blockchain is often presented in a decentralized and distributed implementation, which can be public. Therefore, it can be used as a digital ledger to record transactions across several (potentially many) computers, so that these transactions cannot be altered, at least without needing to tamper all subsequent blocks (therefore, producing a massive alteration that could be easily detected).

Although initially engineered to support the development of a cryptocurrency, throughout the last decade Blockchain has found its way through many different fields of application, both inside and outside the financial world, where the need for immutability and security has thrusted its deployment. Recently, several literature reviews have been published describing some of these fields of application [2], [3], [4], [5], [6].

Additionally, with the introduction of the Ethereum Platform in 2014 [7], the concept of "Smart contracts" arose, as a method of automatically enforcing the execution of the terms of a contract when its conditions are fulfilled by transactions executed in the blockchain.

With the integration of smart contracts, Blockchain technology has enabled companies to consolidate multiple repositories of customer and supplier information into a single data warehouse. Since it uses algorithms to ensure that transactions are valid and authentic, it can help improve processes involving multiple parties. Also, it contains the complete history of all the transactions that have been carried out on the network, along with their time stamp. Therefore, this record can be accessed at any time knowing that this information is fully updated [8].

In any case, although the theoretical foundations of the blockchain are well known, the technology is recent enough as to remain in the peak of innovation, where many new applications of it are being discovered and presented every day. Blockchain improves processes (both in cost and time) as well as allowing hitherto unimaginable scenarios. For the industry, the blockchain enables an improvement in the efficiency and effectiveness of the supply chain. This provides up-to-date and easily accessible documentation on how materials are obtained, purchased, recorded, and used [9].

Due to the origins of the technology, the financial services sector is probably the one that will be able to benefit more quickly from the use of blockchain, by allowing financial transactions between two participants in a safe, reliable and irreversible way. One of the main advantages would come from the supply side, by being able to improve the service catalog. Many of the processes that are currently manual could be automated and speeded up, so new products and services that are currently unfeasible or unmanageable can be offered. Blockchain is a tool that simplifies relationships between large companies and their suppliers, assigning them digital identities that save time and effort by eliminating all the paperwork involved in starting a collaboration between entities, and the introduction of smart contracts can simplify the bureaucratic processes.

In summary, the advantages of the Blockchain technology for the financial sector can be specified in the following aspects:

- Efficiency: Savings in transaction costs, which makes information reconciliation processes more efficient.
- Security: Distributed records allow to verify transactions and collaboration on different nodes ensures their authenticity.
- Transparency: The identities of blockchain users can be cryptographically protected, allowing the system to be completely transparent.
- Accessibility: Can be implemented as a public platform, so any authorized user can obtain a copy of the registry.

In the public administration sector, the possibilities of blockchain technology are enormous. Those responsible for public institutions could ask the following question: How can you create a government on the web that brings institutions closer to citizens? The Blockchain-based public administration model would lead to a decentralized government in which one could even think about the possibility that the vote in an election was done electronically. But to get to this point, it is necessary to implement a series of measures that will compel the citizen to become part of that digital society.

In the real estate field, the possibilities unveiled by the Blockchain have also proven its usefulness by accelerating disruptive processes and rethinking business models. How can a real estate developer use this technology? Two concepts are the key to its implementation: efficiency and elimination of unnecessary intermediaries. If we want to sell a building to a foreign investor, the process is usually very opaque, with many intermediaries, and we do not know the buyer until the last steps of the process. One opportunity is to tokenize assets (with a token being a unit of value issued privately). Each home would then be a token that can be shown to investors so that they have all the information available about the status of the project at all times.

The legal sector is also making progress in taking profit from Blockchain. A particular and original example that can be cited is that of a company in the sector that has resorted to issuing tokens, which in this case are nothing more than hours of legal advice. These tokens are channeled to a foundation that is in charge of working with disadvantaged groups. Subscribers can choose to which wallet they send the tokens and these are destined to the projects of the foundation. In this way, traceability, transparency and visibility of when and how the funds designed to support different projects with social content are used.

Although some of the Blockchain's underlying capabilities provide data confidentiality, standards need to be adopted for organizations using Blockchain in order to protect their organizations from external attacks. In any case, Blockchain technology can be used to protect systems and devices from attacks and improve cybersecurity across industries. Additionally, with a distributed Blockchain, there is no longer a centralized authority controlling the network and verifying the data going through it.

As a result, it should be noted that Blockchain is a technology with enormous possibilities. However, because of it being relatively recent, the regulatory factor has not yet been consolidated. An example of a legal issue that can be formulated regarding the Blockchain is how to make the right to be forgotten compatible with

DOI: 10.9781/ijimai.2020.08.006

a distributed registration technology in which information lives in innumerable nodes? Problems as this arise that may be incompatible with data protection regulations.

More importantly, data stored in Blockchains might constitute a priceless resource for numerous data analysis and intelligence applications. In a scene where machine intelligence relies more and more in huge amounts of data, we cannot underestimate the importance of this information. In some cases, even the Blockchain can be designed so that its inner workings be used to empower an artificial intelligence system [10].

In this special issue, we want to gather some innovative applications that are currently pushing forward the research on Blockchain technologies. In particular, we are interested also in those applications that put the focus on the data, enabling new processes that are able to leverage relevant knowledge from the data.

As stated before, the earliest and most well-known field of application of Blockchain is finance. In this context, it is reasonable to expect that there are some proposals that suggest the hybridization of Blockchain technologies and banking. In this issue, Arjun and Suprabha study the literature to conclude some of the most relevant innovations of Blockchain in banking, as well as the biggest challenges up to date.

Another field of application with a huge potential is healthcare, but it is crucial to put the focus on security and privacy when it comes to storing health-related data from patients. In this issue, Jennath et al. propose a solution based on Blockchain to store health records while addressing privacy concerns, finally outlining the potential of building artificial intelligence models over the e-Health data.

Of course, Blockchain turns out to be a great alternative for those applications requiring to store lots of data that are generated as a result of business-to-business interactions and where some rules must be executed based on certain transactions. This is the case of affiliate systems, where certain fees or deals must take place as a result of established agreements, and the promise of an immutable ledger brings trust to the ecosystem. In this issue, Baldominos et al. present Blockverse, a platform relying on a cloud stored Blockchain that allows tracking in affiliate systems, allowing the computation of advanced analytics and proposing some lines for the construction of intelligent systems relying on the data recorded.

Additionally, Blockchain can also be used to enhance integrity and efficiency on the cloud when it comes to data storage. In particular, this is proved by El Ghazouani et al., who present a cloud storage solution implementing deduplication and auditing mechanisms, while relying in a Blockchain implementation.

But Blockchain is also called to revolutionize other professional fields, such as journalism. In particular, the work by Jurado et al. proposes mechanisms for guaranteeing the traceability of news from their origin, analyzing their evolution and easing fact checking. This work has a potential impact towards improving transparency in the media.

The wideness of application of Blockchain extend even to the countryside, where it can be used to support a pest management system. This use case is described by Lopez et al. in a paper gathered in this special issue, where the architecture of a system for intensive farming is presented, including the ability to detect environmental conditions that might lead to the appearance of pests.

The potential of cross-fertilization of Blockchain and artificial intelligence can also have a direct impact in the security of small and medium enterprises. Such a system is proposed by Lopez et al. in their work, where a platform relying on Blockchain technologies would be in charge of detecting cyberattacks and, if necessary, containing the attack and easing the recovery after the attack takes place.

But Blockchain can also be used beyond banking, enterprises and particulars. As Triana et al. illustrate in their paper, Blockchain along with smart contracts are an extremely useful resource to enhance trust and transparency in the public sector, reducing or preventing corruption and improving efficiency in many bureaucratic procedures.

This special issue will be successful if readers gain a better understanding on how Blockchain can be applied to very diverse areas, and might even be interested in designing, implementing and deploying an innovative solution to a completely different field of knowledge. In such case, the final paper of this SI, published by García-Sáez, can be of special relevance to the avid readers. In this paper, the author put a critical lens on top of Blockchain development and argues the challenges for startups when running a Blockchain solution, posing strategical recommendations to enhance the chance of success.

We hope this Special Issue can provide a better understanding and key insights to readers on how Blockchain and artificial intelligence are cross-fertilizing to revolutionize many aspects in our societies.

A. Baldominos and F. Mochón

References

[1] S. Nakamoto, S., "*Bitcoin: A Peer-to-Peer Electronic Cash System,*" White paper, 2008. Available online: https://bitcoin.org/bitcoin.pdf. (Last visited: 2020/07/08).

[2] F. Casino, T.K. Dasaklis, C. Patsakis, C. "A systematic literature review of blockchain-based applications: Current status, classification and open issues," *Telematics and Informatics*, vol. 36, 2019, pp. 55–81.

[3] M. Xu, X. Chen, G. Kou, "A systematic review of blockchain," *Financial Innovation*, vol. 5, 2019 article no. 27.

[4] M. Conoscenti, A. Vetrò, J.C. De Martin, "Blockchain for the Internet of Things: A systematic literature review," 2016 IEEE/ACS *13th International Conference of Computer Systems and Applications*, IEEE. 2016.

[5] F. Hawlitschek, B. Notheisen, T. Teubner, "The limits of trust-free systems: A literature review on blockchain technology and trust in the sharing economy," *Electronic Commerce Research and Applications*, vol. 29, 2018, pp. 50–63.

[6] Seebacher, S., and Schüritz, R., "Blockchain Technology as an Enabler of Service Systems: A Structured Literature Review," *Lecture Notes in Business Information Processing*, vol. 279, Springer, 2017.

[7] V. Buterin, "*A Next-Generation Smart Contract and Decentralized Application Platform*," White paper, 2014. Available online: https://ethereum.org/en/whitepaper/. (Last visited: 2020/07/08).

[8] M. Iansiti, K.R. Lakhani, "The Truth about Blockchain," *Harvard Business Review*, 2017. Available online: https://hbr.org/2017/01/the-truth-about-blockchain. (Last visited: 2020/08/19).

[9] Consensys, "*Blockchain Use Cases and Applications by Industry*," 2020. Available online: https://consensys.net/blockchain-use-cases/. (Last visited: 2020/08/19).

[10] A. Baldominos, Y. Saez, "Coin.AI: A Proof-of-Useful-Work Scheme for Blockchain-Based Distributed Deep Learning," *Entropy*, vol. 21, no. 8, 2019, article no. 723.

TABLE OF CONTENTS

OPEN ACCESS JOURNAL

ISSN: 1989-1660

The International Journal of Interactive Multimedia and Artificial Intelligence is covered in Clarivate Analytics services and products. Specifically, this publication is indexed and abstracted in: *Science Citation Index Expanded, Journal Citation Reports/ Science Edition, Current Contents®/Engineering Computing and Technology*.

Innovation and Challenges of Blockchain in Banking: A Scientometric View

R. Arjun, K. R. Suprabha *

National Institute of Technology Karnataka, Surathkal Mangalore - 575 025 (India)

Received 30 December 2019 | Accepted 27 January 2020 | Published 28 March 2020

Abstract

Blockchain has been gaining focus in research and development for diverse industries in recent years. Nevertheless, innovations that impact to the banking nurture a potential for disruptive impact globally for economic reasons; however it has received less scholarly attention. Hence the effect of blockchain technologies on banking industry is systematically reviewed. The relevant literature is extracted from Scopus, Web of Science and bibliometric techniques are applied. While a bulk of earlier papers focuses only on bit coins, a broader framework is envisaged that synthesizes interdisciplinary thematic areas for advancement; hence novelty in current work. A few practical and theoretical implications for stakeholders in view of technology, law and management are discussed.

Keywords

Digital Economy, Blockchain, Banking, Scientometrics, Literature Review.

DOI: 10.9781/ijimai.2020.03.004

I. Introduction

In the early 2008, a working proof-of-concept paper was published online being authored under a pseudonym [48]. Slowly the work gained attention of researchers but concepts leading to the invention were published several decades earlier [49]. Put in a nutshell, the paper demonstrated working of bit coin, a type of currency that can be transacted digitally anytime around the world. Within a decade, there has been immense advancements of techniques into big data [28], machine learning [5], internet of money [55], etc. Blockchain, a broader concept that encompasses a model for bit coin transactions, is highly resilient against tampering of the data. By design, this model is "*an open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way*" [31]. A banking entity, traditionally an intermediary between two transacting parties; i.e. buyer and seller of financial asset, has to re-position itself in this new realm. Additionally, risk factor and impact of banking on world economy is well documented [26] [39]. Business intelligence approaches for banking have positive results [30], but IT adoption requires customization for multicultural settings [38]. The digital economy perspective demands sector analysis [6] [34], providing inclusive solutions [32]. Nevertheless, severe dearth of knowledge exists because lack of proper theoretical foundation.

Having demonstrated the impact in finance, technology and government space, the blockchain framework has extended beyond bit coin [64] [65]. A crucial question arises on ways to overcome misconceptions and identify technology caveats for developing world. By such, thin boundaries of hype and reality on blockchain applications get delineated as argued by [53]. Hence, the present study poses 3 empirical research questions: RQ1. *Which are the current technologies within the blockchain ecosystem for competitive advantage to banking institutions*? RQ2. *How will blockchain based digital platforms transform existing value-creating interactions*? RQ3. *What are the implications on legality, technical structure and organizations from an information systems perspective*? The remaining sections of this paper are organized into Section II- background and earlier related work; Section III- methodology and data for research; Section IV- emerging research avenues; Section V- discussion & proposed framework and Section VI- conclusion & suggestions.

* Corresponding author.

E-mail addresses: arjrs123@gmail.com (R. Arjun), suprabha.kr@gmail.com (K. R. Suprabha).

II. Related Work

Basically the systematic review looks on the blockchain research from three lenses. First type were papers focusing exclusively on solutions to banking industry. These include theoretical models or empirical framework with potential for strategic change [3], operational advantages [25], or stock market functions. Second are which highlight practical scope or challenges in platforms, legal, technical and organizational dimensions [7] [24]. Third, specific application oriented papers utilizing conceptual/experimental settings. For core banking based solutions, works on Know-Your-Customer (KYC) [47], bank credit and financing [68], inter-bank payments [43] were explored. For stock market related, lots of innovation is missing. In stock trading business, settlement and clearing process usually consume up to three days (or longer, depending on country and bank types). This implies actual money and shares are frozen for a specific amount of time period. Such delay incurs overhead costs and businesses loss while effects is compounded in high frequency trading scenarios. Major researches have investigated market prediction models [40] [41], decentralization [29] [57], booking registrations [46], etc. Similarly, smart contract is a computer code built into the blockchain to facilitate, verify, or negotiate a contract agreement [21]. Smart contracts operate under a set of pre-set conditions that users agree to. When conditions are met, terms of the agreement are automatically executed, as for insurance [17], etc. Recent works show that open blockchain frameworks like

IBM Fabric process transactions have higher workloads & security than conventional databases [2].

For integrating lessons from banks [27], finance [36], strategy [70], mobile bank [10] and socio-economics [54], the current study use an exploratory approach and inductive reasoning.

III. Methodology and Data

The journal/book articles in Scopus, Web of Science and ACM, IEEE and AIS (Association for Information Systems) conference proceedings published in 2008-2019 (11 years) were collected. Additionally the patents [18], doctoral thesis [56] and industry reports [43] were used as secondary information. Under each category, the articles were chosen for their match with the theme of blockchain and industry focused. In current work, VOSviewer, an open source software tool is used for data analysis and visualization. Specifically methods used to generate the charts are BCAD (Bibliographic Coupling Analysis of Documents) and CCAR (Co-Citation Analysis of References) [66]. Specific aim is to quantitatively assess the research output in peer reviewed journals. From Scopus, 70 documents were retrieved. The Scopus data in CSV file format got downloaded to apply the bibliographic analysis method using VOSviewer software.

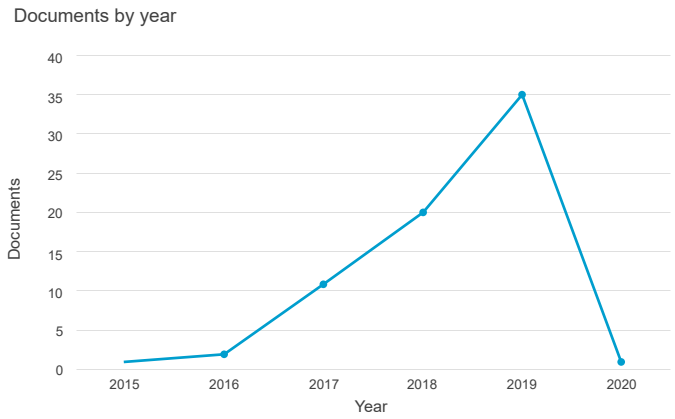


Fig. 1. Research trend (source: Scopus).

Scopus query format: *(TITLE-ABS-KEY (blockchain) AND TITLE-ABS KEY (*bank)) AND (LIMIT-TO (SRCTYPE,"j"))).*

This query in database selects research articles having title, abstract or keywords as "*blockchain*" and "*bank*". Articles with 1 citation and published during 2008-2019 were used. Fig. 1 shows renewed research interest in blockchain. Only less than 5 articles were published in 2015 but by 2019, no. increased to 35 (85 % increase).

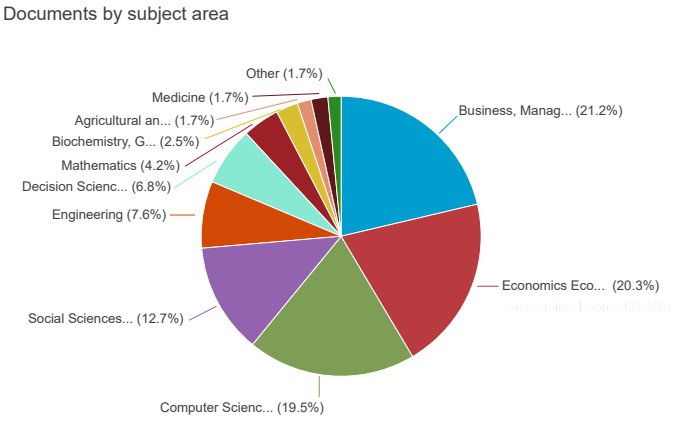


Fig. 2. Fields of research (source: Scopus).

As seen in Fig. 2, the majority of research studies constituted from Business & Management (21.2%), followed by Economics and related fields (20.3%), Computer science (19.5%) & Social sciences (12.7%).

Other areas contributed low to research efforts. Interestingly blockchain applications also find use in biochemistry (2.5%), agriculture and medicine (1.7% each) and others [61].

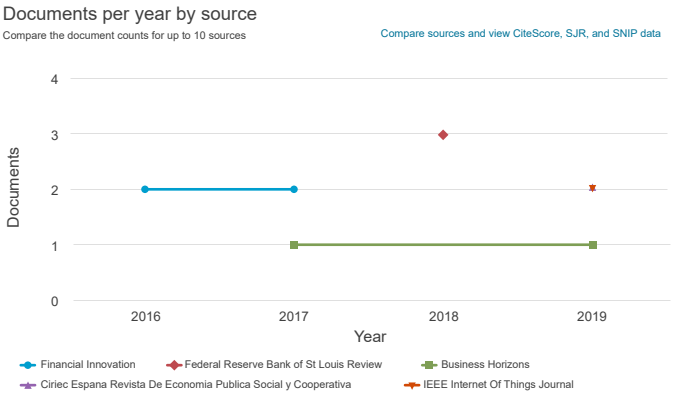


Fig. 3. Publication outlets (source: Scopus).

Fig. 3 shows the major publication peer reviewed journals that encourage blockchain research. Financial innovation has articles from 2016 to present, followed by Federal Reserve Bank of St. Louis, Business horizons and IEEE.

CCAR (Co-Citation Analysis of References): CCAR method is forward looking assessment of past works. Here the frequency of citation received for two published articles together by another paper gets counted. This method analyzes from past research perspective to highlight an intellectual foundation i.e., references highly co-cited, recognizing seminal studies for field.

Fig. 4. CCAR output (source: VOSviewer).

As in Fig. 4, all studies were processed to visualize cited articles. The methodical assumption is that results support/contradict a particular thought school and evolve overtime. Here Nakamoto (2008) [48] appeared twice in visualization since this article is not a formally published study but rather a proof-of-concept. The other influential articles were published in 2017 [31] by HBR and an IEEE IoT study [12]. Study [31] used non-expert language to outline technical foundations of blockchain. They dispel the hype by highlighting pros and cons of blockchain systems. [12] designed framework enabling operation of blockchain + IoT devices. To find the studies on banking with stock market impact, a second search is done after varying the parameters.

BCAD (Bibliographic Coupling Analysis of Documents): This method helps in recognizing present trend/direction of research efforts. Here the measure of similarity is found among studies. This is done by identifying common references between two distinct research works.

Basically, the procedure selects any research article having title, abstract or keywords as "*blockchain*" logical AND "*bank*" or "*stock market*" terms. From this, around 217 results were retrieved. Out of this, studies cited at least 1 are chosen by setting threshold value. Following this, 85 items were filtered. Out of this, 52 journal items are grouped into 8 clusters by VOSviewer (Fig. 5).

Scopus query syntax: *(TITLE-ABS-KEY (blockchain) AND TITLE-ABS-KEY ("stock market") OR TITLE-ABS-KEY (bank))*

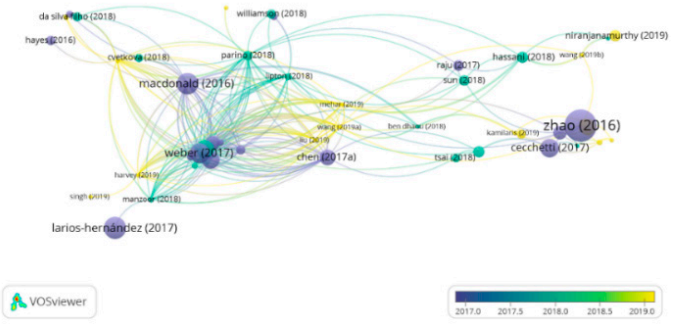


Fig. 5. BCAD output (source: VOSviewer).

The major studies are summarized in Table I in Appendix. The top Scopus paper [25] had prominence percentile score of 99.945, Field weighted citation impact 9.51 (Elsevier based metrics data).

Web of Science Analysis: Before executing WoS database query, under the databases dropdown select options, "Web of Science core collection" is chosen. After this process, 28 results are obtained, from which 22 are articles, 4 review papers, 2 book chapters and editorial material and 1 proceedings paper. From these, articles with 1 citation are chosen. After this, 17 items were visualized into 4 clusters (Fig. 6). Some of the research works that already appeared in Scopus were omitted to avoid duplication.

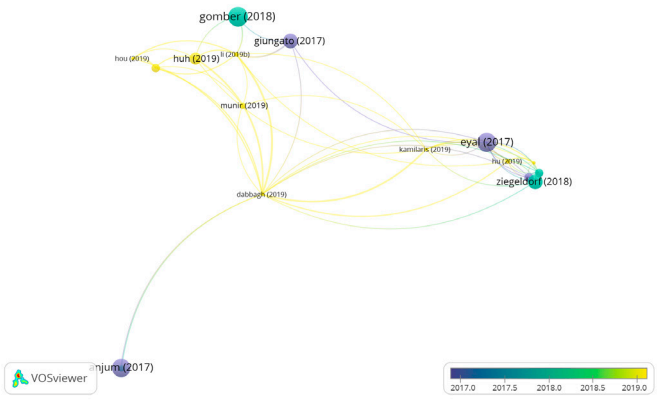


Fig. 6. BCAD output (source: VOSviewer).

Query structure*: TOPIC: (blockchain) AND TOPIC: (bank) Timespan: All years. Indexes: SCI-EXPANDED.*

WoS articles were not analyzed in-depth in the current study as earlier works explored the same [14]. Other works had looked to financial technology (Fintech) comprising services as well [44].

Citation Analysis: Fig. 7 shows the citation analysis of research published in Web of Science. Here the output from 89 records for Total Citing Articles without self-citations are given. It is interesting that citation trend had least quantum of links by materials science and medicinal fields. A few studies and industry reports indicate that manufacturing industry sector lacks significant advances of blockchain where operations research has disruptive potential [4].

Query structure: *TOPIC: (blockchain) AND TOPIC: (bank)* is visualized under the Results Analysis page of WoS database interface.

Fig. 7. Citation analysis (source: WoS).

Table II in Appendix lists the major works. Top WoS study [23] had average citations/year 9.33 as of Jan-2020 (Clarivate Analytics data).

IV. Emerging Areas of Research

From review, the major themes are **banking**, **information systems**, **innovation**, **law**, **finance** [63], **sustainability**, **entrepreneurship** & **digital infrastructure (standards, scalability** [72], **security**, & **privacy** [67]). This section further examines the current directions like:

Social media leveraged prediction: Interesting research directions are: *What are prediction methods to forecast future market conditions utilizing unstructured data sources*? Research groups globally have increased attention to capture user generated content (UGC), social media for advanced explanatory/predictive models [56] [59].

Business model transformations: Future research questions of interest can be: *What are the innovative business models that combine platform based technologies and incentivized interactions*? With empirical evidence of positive correlation between bit coin transaction volume and human development, if technology solutions affect Human Development Index (HDI) in developing economies, this leads to sustainable outcomes [58] [18].

Machine learning based service evolution: Future research can investigate areas like: How do machine learning models fit into the ethical or legal framework of blockchain system? Ex: *Can government/ industry body inherently frame legal implications of data privacy/ security (GDPR) and similar data protection rules into secured blockchain*? Findings from these ideas can pave way for solutions to ethics, machine learning, and security researcher's community [62].

Market reaction for blockchain: Looks variations in investor reaction on short-term and long-term basis for information disclosure of bit coin prices by firms. Study compared results with SEC (Securities and Exchange Council), USA concerns and local media reports. Future research could investigate if managers exploit this momentary investor overreaction, sell stock, engaging insider trading activities, exercising options, etc. [11].

Digital platforms as infrastructure: An editorial note [13] points interesting avenues of research study that simplify intellectual property rights (IPR) to create ownership, protect rights using technologies like blockchains. Fostering the debate to balance the goals of privacy against the desire for economic growth is needed. *Will social media be viewed as "critical infrastructure" given their ability to influence critical societal functions even as elections*? If so, what are appropriate regulatory regimes? Which is the suitable role for government intervention into how platforms operate? Is self-regulation in digital platforms feasible? Ex. Libra from Facebook proposed as blockchain based open source digital currency.

V. Discussion and Proposed Framework

Broadly, the issues associated in blockchain technologies specific to banking were identified. Major drawbacks of blockchain or crypto currencies are technology cost incurred, low transaction throughput, illegal use like terrorism, drug trade, cyber-crimes, etc. [20]. Therefore, need arises for industry consensus to solve challenges with collective benefit [60].

Regulatory landscape: The legality of blockchain and bit coin is subject of intense public confusion and government scrutiny, hence still a grey area. Adding to complexity, there exists ambiguity at multiple countries [33]. For example, in India, where services including banking contribute more than 50% GDP (2018), it was announced in 2018 that bit coin was not a legal tender for transactions, but blockchain technology is promoted in payment systems.

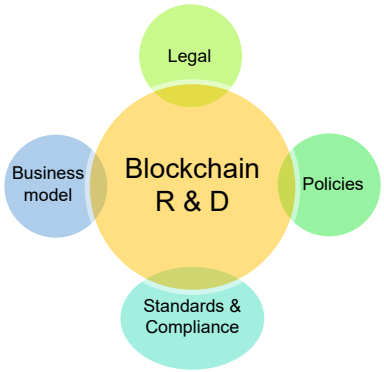


Fig. 8. Blueprint for Blockchain R&D.

Fig. 8 depicts the conceptual map of regulatory aspects surrounding blockchain. As seen in previous ex: the logical connective in legal-economic policy-compliance and business model emerges as paramount [45]. Extent of controls, under/over regulation, policies in turn affect the ecosystem. The ambit of policies must be delineated in R&D and practice by a modular approach.

Infrastructure requirements: Whenever scope exists for security, privacy, technical capability, organizations in blockchain ecosystem must assess their SWOT [50]. This process can foster better Human Computer Interaction (HCI) modules [16]. Coordinating these well helps investor's perspective of solution of business model. Fig. 9 depicts hierarchy required by generic business/non-profit organizations venturing in blockchain.

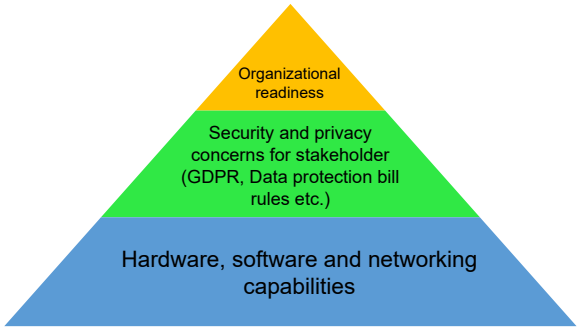


Fig. 9. Requirements hirearchy.

Socio-economic drivers: The blockchain ecosystem, like any other major technological up gradation or shifting, requires a multilevel effort and time. The process of technology use and adoption of literature and models, such as TAM (Technology Acceptance Model)/UTAT (Unified Theory of Acceptance and Use of Technology), and status quo theory finds relevance in this context [9].

Another interesting study analyzed the changes brought by bit coin from bank currency notes even culturally [24]. The framework attempts in coherent logic linking theories from information systems, management and sociology.

Effect of bit coin prices and impact of alternate bank systems in economy policy had been investigated earlier [15] [35] [52] denoting economic impact. Fig. 10 outlines a roadmap to develop blockchain systems in cross-country basis. It synthesizes earlier work [54], to map concerns [19], pitfalls of regulation, technical/organization aspects.

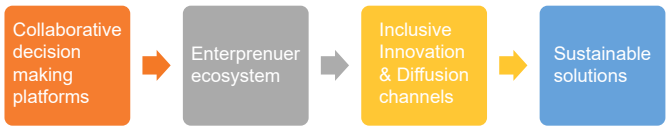


Fig. 10. Socio-economic factors.

VI. Conclusion

One limitation of this study is that only 72 research works could be covered in the systematic review. The external databases (JSTOR etc.) hold more literature but is outside the scope of this paper. Also, citation data from Scopus/WoS compared to Google Scholar or reference managers like Mendeley, etc. varies highly. Hence the criteria to choose research is only judgmental; for ex: citations count as proxy measure for impact. While earlier works limited to WoS, present study overcomes with Scopus also. There are two papers visually located far from clusters centroids, indicating lack of extended work on entrepreneurship and standards that hinders blockchain [37] [1]. Also, general inference is that blockchain technology is affected by government, infrastructure and organizational dynamics. While Scopus has literature addressing specific applications and empirical practice, WoS offers insights into the theoretical dimensions and societal implication aspects.

Now answering the posed questions, RQ1: In developing economies, consortium led blockchain systems like R3 and strategizing of Fintech 2.0 roadmap are essential. Instead of replicating strategies of firms like Google, Amazon, Apple or Facebook, etc. the digital banking systems must be redesigned for newer modalities. One such example is Paytm of India. Most of these firms are developing in-house payments ex: Apple Pay, Google Pay enabled with near field technologies (NFC), contactless payments or facial recognitions. Hence, a paradigm shift to evolve banking models on services and operations front is key to stay relevant. RQ2 answer: Synergy of banking management information systems with worldwide real-time inter-bank network is the future. By monitoring global risks, functionality is augmented using interactive and business intelligence tools for mobile banking, stock market predictions, credit evaluations, risk management, etc. for investors and customers. Banking-as-a-service (BaaS) has evolved into service platform impacting the customer-businesses-government interactions. RQ3 answer: The standardization in transactions, privacy/security infrastructure and resolution mechanism by large banking institutions can influence smaller and non-banking public and private blockchain entities. New corporate social media and customer relations management will hence be strong catalysts for operations strategy. The government, industry and stakeholder community must undertake holistic & co-creation approach for blockchain solutions. The current paper intends as blueprint to transform concerns and formulate actionable deliverables. Future work can adopt causal research using either panel data or survey based investigation into effectiveness of blockchain solutions.

Appendix

TABLE I. Top 10 Impact Making Research (Based on Citations)

Author/Study	Objective & Methods	Findings and limitations	Major theme and suggestions.	Total citations (TC)
Guo and Liang (2016) [25]	To assess the blockchain for banking industry in the Chinese context. Uses primary data and industry reports.	Blockchain + banks (FinTech 2.0) has superior customer experience, efficiency, cost and security.	**Banking**. Shift to R3, consortium led blockchain. Payment clearing system, bank credit information systems are vital.	91
Zhao *et al.* (2016) [71]	*Editorial note*. Introduction to 7 academic papers in special edition.	Multidisciplinary nature of blockchain research. WoS and SSRN articles grown from 0 in 2014 to 11 and 79 respectively by 2016.	**Innovation**. Challenges remain on theoretical issues of societal impact, smart contracts, security etc.	57
Kiviat (2015) [33]	To integrate research from diverse domains and educate legal community, help practitioners. Primary and secondary data, legal proceedings analysis.	Regulatory patchworks are underway in USA but virtual currencies pose uncertainty.	**Legal.** Policymaker must be cautious and have precision in tailoring the scope of regulation	44
Nguyen (2016) [51]	To study role of blockchain in sustainable outcomes investigating in finance industry. Secondary sources and data used.	Limits competition as blockchain network will be shared. Risks from payment, technology cost and stability hinder.	**Sustainability**. Further study on crowdsourcing, payment system. Banks face pressure from investors and competing MNC's.	33
Larios-Hernández (2017) [37]	To explain lack of formal bank account. Leverage the Blockchain based financial inclusion. Uses fuzzy-set Qualitative Comparative Analysis (fsQCA)	Variety of supply-and demand-related causal factors. Informal peer-to-peer practice is customary.	**Entrepreneurship/ financial inclusion**. Business logic of existing financial institutions hinder the solutions for unbanked.	27
MacDonald *et al.* (2016) [42]	To explore the implications of blockchain technology on the future of banking. Uses systematic review and theory building methods.	Enables to competing banks organizations, shifting banking transactions to out of centralized organizations and into decentralized markets.	**Economy**. Need to find balance between technology adoption and governance.	25
Weber *et al.* (2017) [69]	To find out limitations of two mainstream blockchains, *Ethereum* and *Bit coin*. Uses simulation experiments.	Network reordering impact transaction commit times and even counter balance the effects of transaction fees and gas price.	**Distributed systems**. Application developers must consider factors that may dramatically affect quality of service.	23
Giungato *et al.* (2017) [22]	To define and evaluate sustainability of bit coin, and environmental impacts, social issues and economic aspects.	Transition of the entire monetary system to a new crypto currency can result in an undesirable amount of energy consumed to mine new bit coins.	**Sustainability**. Extend to medical data, energy generation and distribution in micro-grids at the citizen level, block-stack & state-driven crypto currencies.	20
Cecchetti *et al.* (2017) [8]	To design *Solidus*, a protocol for confidential transactions on public blockchains, such as those required for asset transfers with on-chain settlement.	Implements Solidus and present a set of benchmarks indicating that the system is efficient.	**Data privacy**. Solidus hides transaction values and transaction graph (i.e., identities of transacting entities) while maintaining the public verifiability.	20
Berentsen and Schär (2018) [7]	An introduction to bit coin, operation and applications.	Price volatility and scaling issues makes bit coin adoption a complex affair.	**Crypto currencies.** Bit coin and crypto assets have potential to become asset class.	18

Table II. Top 5 Impact Making Research (Based on Citations)

Author/Study	Objective & Methods	Findings and limitations	Major theme and suggestions.	Total citations(TC)
Gomber *et al*. 2018 [23]	To review the technology innovation, process disruption, & services transformations. Quantitative & qualitative data.	Assessing the factors outlined in study and discussed needs to be repeated.	**Information systems**. Scope for developing research by interdisciplinary sources, designs, theory and methodologies.	28
Eyal (2017) [17]	To gain overview of protocols, distributed-ledger technology (DLT). Exploration of boundaries of blockchain beyond bit coin.	Requirements of blockchains for crypto currencies with that of FinTech vary drastically— Ex: transaction throughput, security and privacy. Clear distinction of FinTech and crypto currencies.	**Financial applications**. Direct and effective collaboration required between the FinTech industry and the blockchain scientific engineering community.	25
Anjum *et al*. 2017 [1]	To compare the types and performance of blockchain architectures.	Platforms for specific sectors and application domains are emerging. Current models exhibit limitations in areas related to scalability, flexibility, and governance.	**Standards**. Standardization activity required to as technologies enablers and for interoperability.	23
Treleaven *et al*. 2017 [63]	*Editorial note*. On special issue covering journal research articles.	As blockchain technology evolves, it becomes disruptive for other technologies such as big data, the IoT, intelligent assistants, and autonomous vehicles with opportunities and unintended social consequences.	**Multi-themed.** Smart contracts could become the management framework for many of private records in the future.	21
Ziegeldorf *et al*. 2018 [72]	To study *CoinParty*, an efficient decentralized mixing service for users to reestablish financial privacy.	The prototype implementation scales to large no. of users, achieves data anonymity higher level than earlier models.	**Data privacy**. Third party independent deployment feasible for organizations.	15

Acknowledgment

The first author is grateful for research scholarship support from National Institute of Technology Karnataka. Both the authors express invaluable appreciation to the reviewers, editorial team for suggesting improvements in the paper. Authors declare no conflict of interests.

References

[1] Anjum, A., Sporny, M. and Sill, A., 2017. Blockchain standards for compliance and trust. *IEEE Cloud Computing*, *4*(4), pp.84-90.

[2] Ankur Sharma, Felix Martin Schuhknecht, Divya Agrawal, and Jens Dittrich. 2019. Blurring the Lines between Blockchains and Database Systems: the Case of Hyperledger Fabric. In *Proceedings of the 2019 International Conference on Management of Data* (*SIGMOD '19*). ACM, 105–122. DOI: https://doi.org/10.1145/3299869.3319883

[3] Ashta, A. and Biot-Paquerot, G., 2018. Fintech evolution: Strategic value management issues in a fast changing industry. *Strategic Change*, *27*(4), pp.301-311.

[4] Babich, V. and Hilary, G., 2019. Distributed ledgers and operations: What operations management researchers should know about blockchain technology. *Manufacturing & Service Operations Management*. https://doi.org/10.1287/msom.2018.0752

[5] Baldominos, A., Saez, Y., 2019. Coin.AI: A Proof-of-Useful-Work Scheme for Blockchain-Based Distributed Deep Learning. *Entropy*, *21*, p.723.

[6] Benito, S., de Juan, R., Gómez, R. and Mochón, F., 2015. Differences in Measuring Market Risk in Four Subsectors of the Digital Economy. *International Journal of Interactive Multimedia and Artificial Intelligence*, *3*(2), pp.9-16.

[7] Berentsen, A. and Schär, F., 2018. A Short Introduction to the World of Cryptocurrencies. *Federal Reserve Bank of St. Louis Review*, *100*(1), pp.1-16.

[8] Cecchetti, E., Zhang, F., Ji, Y., Kosba, A., Juels, A. and Shi, E., 2017, October. Solidus: Confidential distributed ledger transactions via PVORM. In *Proc. of 2017 ACM SIGSAC Conference on Computer and Communications Security* (pp. 701-717). ACM.

[9] Chatfield, A.T. and Reddick, C., 2019. Blockchain Investment Decision Making in Central Banks: A Status Quo Bias Theory Perspective. AMCIS 2019 Proceedings. https://aisel.aisnet.org/amcis2019/digital_government/digital_government/3/

[10] Chen, Z., Li, Y., Wu, Y. and Luo, J., 2017. The transition from traditional banking to mobile internet finance: an organizational innovation perspective-a comparative study of Citibank and ICBC. *Financial Innovation*, *3*(1), p.12.

[11] Cheng, S.F., De Franco, G., Jiang, H. and Lin, P., 2019. Riding the Blockchain Mania: Public Firms' Speculative 8-K Disclosures. *Management Science* https://doi.org/10.1287/mnsc.2019.3357

[12] Christidis, K. and Devetsikiotis, M., 2016. Blockchains and smart contracts for the internet of things. *IEEE Access*, *4*, pp. 2292-2303.

[13] Constantinides, P., Henfridsson, O. and Parker, G.G., 2018. Introduction-Platforms and Infrastructures in the Digital Age. *Information Systems Research*, *29*(2), pp.381-400.

[14] Dabbagh, M., Sookhak, M. and Safa, N.S., 2019. The evolution of blockchain: A bibliometric study. *IEEE Access*, *7*, pp.19212-19221.

[15] Dwyer, G.P., 2015. The economics of Bit coin and similar private digital currencies. *Journal of Financial Stability*, *17*, pp.81-91.

[16] Elsden, C., Manohar, A., Briggs, J., Harding, M., Speed, C. and Vines, J., 2018. Making sense of blockchain applications: A typology for HCI. In *Proc. of the 2018 CHI Conference on Human Factors in Computing Systems* (p. 458). ACM.

[17] Eyal, I., 2017. Blockchain technology: Transforming libertarian cryptocurrency dreams to finance and banking realities. *Computer*, *50*(9), pp.38-49.

[18] Fay, T. and Paniscotti, D., Nasdaq Inc, 2016. *Systems and methods of blockchain transaction recordation*. U.S. Patent Application 15/086,801.

[19] Fernandez-Vazquez, S., Rosillo, R., De La Fuente, D. and Priore, P., 2019. Blockchain in FinTech: A mapping study. *Sustainability*, *11*(22), p.6366.

[20] Foley, S., Karlsen, J.R. and Putniņš, T.J., 2019. Sex, drugs, and bit coin: How much illegal activity is financed through cryptocurrencies?. *The Review of Financial Studies*, *32*(5), pp.1798-1853.

[21] Gatteschi, V., Lamberti, F., Demartini, C., Pranteda, C. and Santamaría, V., 2018. Blockchain and smart contracts for insurance: Is the technology mature enough?. *Future Internet*, *10*(2), p.20.

[22] Giungato, P., Rana, R., Tarabella, A. and Tricase, C., 2017. Current trends in sustainability of bit coins and related blockchain technology. *Sustainability*, *9*(12), p.2214.

[23] Gomber, P., Kauffman, R.J., Parker, C. and Weber, B.W., 2018. On the fintech revolution: interpreting the forces of innovation, disruption, and transformation in financial services. *Journal of Management Information Systems*, *35*(1), pp.220-265.

[24] Grossman, J.H., 2019. Passing cash from bank notes to bit coin: standardizing money. *Journal of Cultural Economy*, *12*(4), pp.299-316.

[25] Guo, Y. and Liang, C., 2016. Blockchain application and outlook in the banking industry. *Financial Innovation*, *2*(1), p.24.

[26] Haldane, A.G. and May, R.M., 2011. Systemic risk in banking ecosystems. *Nature*, *469*(7330), p.351. https://doi.org/10.1038/nature09659

[27] Harris, W. and Wonglimpiyarat, J., 2019, Blockchain platform and future bank competition, Foresight. https://doi.org/10.1108/FS-12-2018-0113

[28] Hassani, H., Huang, X. and Silva, E., 2018. Banking with blockchain-ed big data. *Journal of Management Analytics*, *5*(4), pp.256-275.

[29] Hayes, A., 2016. Decentralized banking: monetary technocracy in the digital age. In *Banking Beyond Banks and Money* (pp. 121-131). Springer, Cham.

[30] Hu, D., Zhao, J.L., Hua, Z. and Wong, M., 2012. Network-Based Modeling and Analysis of Systemic Risk in Banking Systems. *MIS Quarterly*, *36*(4).

[31] Iansiti, M. and Lakhani, K.R., 2017. The truth about blockchain. *Harvard Business Review*, *95*(1), pp.118-127.

[32] Kehr, H., Tonkin, G. and Bihler, R., 2017. The Unbanked Don't Need More Brick and Mortar Banks. In *Shaping the Digital Enterprise* (pp. 139-156). Springer, Cham.

[33] Kiviat, T.I., 2015. Beyond bit coin: Issues in regulating blockchain transactions. *Duke LJ*, *65*, p.569.

[34] Kou, G., Chao, X., Peng, Y., Alsaadi, F.E. and Herrera-Viedma, E., 2019. Machine learning methods for systemic risk analysis in financial sectors. *Technological and Economic Development of Economy*, pp.1-27.

[35] Krainer, R.E., 2017. Economic stability under alternative banking systems: Theory and policy. *Journal of Financial Stability*, *31*, pp.107-118.

[36] Kruglova, I.A. and Dolbezhkin, V.A., Objective Barriers to the Implementation of Blockchain Technology in the Financial Sector. In *2018 International Conference on Artificial Intelligence Applications and Innovations (IC-AIAI)* (pp. 47-50). IEEE.

[37] Larios-Hernández, G.J., 2017. Blockchain entrepreneurship opportunity in the practices of the unbanked. *Business Horizons*, *60*(6), pp.865-874.

[38] Leonardi, P.M., Bailey, D.E., Diniz, E.H., Sholler, D. and Nardi, B.A., 2016. Multiplex Appropriation in Complex Systems Implementation: The Case of Brazil's Correspondent Banking System. *MIS Quarterly*, *40*(2), pp.461-473.

[39] Lipton, A., 2018. Blockchains and distributed ledgers in retrospective and perspective. *The Journal of Risk Finance*, *19*(1), pp. 4-25.

[40] Liu, X. and Lin, N., 2018. An Automatic Discovery Process of Stock Value Information with Software Industry Based on Blockchain. In *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)* (pp. 136-139). IEEE.

[41] Liu, X. and Yu, T., 2018. An automatic pattern recognition value system with listed banks based on blockchain. In *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)* (pp. 1850-1854).

[42] MacDonald, T.J., Allen, D.W. and Potts, J., 2016. Blockchains and the boundaries of self-organized economies: Predictions for the future of banking. In *Banking beyond banks and money* (pp. 279-296). Springer, Cham. https://doi.org/10.1007/978-3-319-42448-4_14

[43] Meszaros, R., Adachi, D., Dharamsi, H., Yetiskin, B., & Thomas, P. (2016). Blockchain Technology: How banks are building a real-time global payment network. *Accenture Mobility*. https://www.accenture.com/_acnmedia/pdf-35/accenture-blockchain-how-banks-building-real-time-global-payment-network.pdf Accessed on 18-11-2019

[44] Milian, E.Z., Spinola, M.D.M. and de Carvalho, M.M., 2019. Fintechs: A literature review and research agenda. *Electronic Commerce Research and Applications*, *34*, p.100833.

[45] Mik, E., 2017. Smart contracts: terminology, technical limitations and real world complexity. *Law, Innovation and Technology*, *9*(2), pp.269-300.

[46] Miraz, M.H. and Donald, D.C., 2018. Application of Blockchain in Booking and Registration Systems of Securities Exchanges. In *2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE)* (pp. 35-40). IEEE.

[47] Moyano, J.P. and Ross, O., 2017. KYC optimization using distributed ledger technology. *Business & Information Systems Engineering*, *59*(6), pp.411-423.

[48] Nakamoto, S. Bit coin: A Peer-to-Peer Electronic Cash System. 2008. Available online: https://bit coin.org/bit coin.pdf Accessed on 30-09-2019.

[49] Narayanan, A. and Clark, J., 2017. Bit coin's academic pedigree. *Communications of the ACM*, *60*(12), pp.36-45.

[50] Niranjanamurthy, M., Nithya, B.N. and Jagannatha, S., 2018. Analysis of Blockchain technology: pros, cons and SWOT. *Cluster Computing*, pp.1-15.

[51] Nguyen, Q.K., 2016. Blockchain-a financial technology for future sustainable development. In *2016 3rd International Conference on Green Technology and Sustainable Development (GTSD)* (pp. 51-54). IEEE.

[52] Othman, A.H.A., Alhabshi, S.M., Kassim, S. and Sharofiddin, A., 2019. The impact of cryptocurrencies market development on banks' deposits variability in the GCC region. *Journal of Financial Economic Policy*. https://doi.org/10.1108/JFEP-02-2019-0036

[53] Peck, M.E., 2017. Blockchain world-Do you need a blockchain? This chart will tell you if the technology can solve your problem. *IEEE Spectrum*, *54*(10), pp.38-60.

[54] Parino, F., Beiró, M.G. and Gauvin, L., 2018. Analysis of the Bit coin blockchain: socio-economic factors behind the adoption. *EPJ Data Science*, *7*(1), p.38.

[55] Peters, G.W. and Panayi, E., 2016. Understanding modern banking ledgers through blockchain technologies: Future of transaction processing and smart contracts on the internet of money. In *Banking beyond banks and money* (pp. 239-278). Springer, Cham.

[56] Phillips, R.C., 2019. *The Predictive Power of Social Media within Cryptocurrency Markets* (Doctoral dissertation, UCL (University College London). https://discovery.ucl.ac.uk/id/eprint/10071683/

[57] Pop, C., Pop, C., Marcel, A., Vesa, A., Petrican, T., Cioara, T., Anghel, I. and Salomie, I., 2018. Decentralizing the stock exchange using blockchain an ethereum-based implementation of the Bucharest Stock Exchange. In *2018 IEEE 14th International Conf. on Intelligent Computer Communication and Processing (ICCP)* (pp. 459-66).

[58] Qureshi, S. and Xiong, J., 2019. The effect of Bit coin Transactions on Human Development: Emerging Business Models. 25th Americas Conference on Information Systems, Cancun.https://aisel.aisnet.org/amcis2019/global_dev/global_dev/10/

[59] Rousidis, D., Koukaras, P. and Tjortjis, C., 2019. Social media prediction: a literature review. *Multimedia Tools and Applications*, pp.1-33. https://doi.org/10.1007/s11042-019-08291-9

[60] Salman, T., Jain, R. and Gupta, L., 2018. Probabilistic Blockchains: A Blockchain Paradigm for Collaborative Decision-Making. In *2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)* (pp. 457-465)

[61] Smetana, S., Seebold, C. and Heinz, V., 2018. Neural network, blockchain, and modular complex system: The evolution of cyber-physical systems for material flow analysis and life cycle assessment. *Resources, Conservation and Recycling*, *133*, pp.229-230.

[62] Sun Yin, H.H., Langenheldt, K., Harlev, M., Mukkamala, R.R. and Vatrapu, R., 2019. Regulating cryptocurrencies: a supervised machine learning approach to de-anonymizing the bit coin blockchain. *Journal of Management Information Systems*, *36*(1), pp.37-73.

[63] Treleaven, P., Brown, R.G. and Yang, D., 2017. Blockchain technology in finance. *Computer*, *50*(9), pp.14-17.

[64] Tschorsch, F. and Scheuermann, B., 2016. Bit coin and beyond: A technical survey on decentralized digital currencies. *IEEE Communications Surveys & Tutorials*, *18*(3), pp.2084-2123.

[65] Underwood, S., 2016. Blockchain beyond bit coin. *Communications of the*

ACM, *59*(11), pp.15-17.

[66] Van Eck, N. and Waltman, L., 2009. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), pp.523-538.

[67] Wang, H., Ma, S., Dai, H.N., Imran, M. and Wang, T., 2019. Blockchain-based data privacy management with nudge theory in open banking. *Future Generation Computer Systems*. https://doi.org/10.1016/j.future.2019.09.010

[68] Wang, R., Lin, Z. and Luo, H., 2019. Blockchain, bank credit and SME financing. *Quality & Quantity*, *53*(3), pp.1127-40.

[69] Weber, I., Gramoli, V., Ponomarev, A., Staples, M., Holz, R., Tran, A.B. and Rimba, P., 2017. On availability for blockchain-based systems. In *2017 IEEE 36th Symposium on Reliable Distributed Systems (SRDS)* (pp. 64-73). IEEE.

[70] Woodside, J.M., Augustine Jr, F.K. and Giberson, W., 2017. Blockchain technology adoption status and strategies. *Journal of International Technology and Information Management*, *26*(2), pp.65-93.

[71] Zhao, J.L., Fan, S. and Yan, J., 2016. Overview of business innovations and research opportunities in blockchain and introduction to the special issue. *Financial Innovation*, *2*(1), pp.1-7.

[72] Ziegeldorf, J.H., Matzutt, R., Henze, M., Grossmann, F. and Wehrle, K., 2018. Secure and anonymous decentralized Bit coin mixing. *Future Generation Computer Systems*, *80*, pp.448-466.

Arjun R

Arjun R is a doctoral scholar specialized in Information systems at National Institute of Technology Karnataka, India. He has Masters in Software Engineering from Cochin University of Science & Technology. Research interests are Information Systems & Cybernetics. He has published research in international journals and reputed conferences.

Suprabha K.R

Dr. Suprabha K.R is an Assistant Professor at the NIT Karnataka, India. She has obtained PhD from VTU, India. She has guided research scholars and published in reputed journals and conferences while winning best paper awards. She is an investigator for funded research projects.

Blockchain for Healthcare: Securing Patient Data and Enabling Trusted Artificial Intelligence

H.S. Jennath[1]*, V.S. Anoop[2], S. Asharaf[3]

[1] Data Engineering Lab, Indian Institute of Information Technology and Management – Kerala (IIITM-K), Thiruvananthapuram, 695581 (India)
[2] Kerala Blockchain Academy (KBA), Indian Institute of Information Technology and Management – Kerala (IIITM-K), Thiruvananthapuram, 695581 (India)
[3] Indian Institute of Information Technology and Management – Kerala (IIITM-K), Thiruvananthapuram, 695581 (India)

Received 10 December 2019 | Accepted 11 January 2020 | Published 2 July 2020

Abstract

Advances in information technology are digitizing the healthcare domain with the aim of improved medical services, diagnostics, continuous monitoring using wearables, etc., at reduced costs. This digitization improves the ease of computation, storage and access of medical records which enables better treatment experiences for patients. However, it comes with a risk of cyber attacks and security and privacy concerns on this digital data. In this work, we propose a Blockchain based solution for healthcare records to address the security and privacy concerns which are currently not present in existing e-Health systems. This work also explores the potential of building trusted Artificial Intelligence models over Blockchain in e-Health, where a transparent platform for consent-based data sharing is designed. Provenance of the consent of individuals and traceability of data sources used for building and training the AI model is captured in an immutable distributed data store. The audit trail of the data access captured using Blockchain provides the data owner to understand the exposure of the data. It also helps the user to understand the revenue models that could be built on top of this framework for commercial data sharing to build trusted AI models.

Keywords

Blockchain, Healthcare, Data Privacy And Security, Trusted Artificial Intelligence, Interoperability, Patient-Centric Data Management, Provenance, Traceability.

DOI: 10.9781/ijimai.2020.07.002

I. Introduction

HEALTHCARE data is very sensitive due to the inclusion of personal information, hence this industry demands requirements like security and privacy of storage and access of data, confining to legal obligations in protecting their health information [1]. Advances in information technology(IT) enabled the digitization of the health records which further made the data sharing across the stakeholders much easier. Traditionally healthcare datarests within the vicinity of the healthcare provider. These organizations usually manage their in house data of their patients using healthcare customer relationship management systems or electronic health records(EHR). EHR permits digital data capture as well as storage of medical records and access to authorized users [15]. EHRs can assist doctors or healthcare professionals with decision making and can provide insight into the way a patient is treated. Recent accelerated digitization adoption in healthcare aims to provide better patient care, with more accurate analysis and diagnosis and secured access to medical data on demand at a reduced cost. This is achieved by avoiding multiple diagnoses or redundant drug administration, etc. With the advent of Artificial Intelligence (AI), the momentum of innovation in digitization in healthcare is gaining momentum and the industry is now all set to accelerate further with Blockchain technology [18][19].

* Corresponding author.
E-mail address: jennath.res16@iiitmk.ac.in

Advances in the area of wireless communication, distributed storage and edge computing, paved the path to accelerated development and deployments of wearable devices and cloud storage for ubiquitous monitoring and recording of health-related information. The sensitive data stored and processed in e-health systems are mandated to have adopted enough security and privacy measures to prevent disclosure or leakage of medical data as it causes a significant impact on a patient's social life. Leaked medical info would cause a lot of problems in patients personal life such as a loan request rejection, surge in insurance premium, rejection of job application, etc. Hence any deliberate or accidental data breach of healthcare records violating the privacy or confidentiality concerns of patients would lead to severely penalizing the healthcare provider or the IT (Information Technology) infrastructure provider [16], [17].

Mobile Health (mHealth) is another prominent field in the healthcare industry that employs IoT devices or wearable sensors, low-power body-area wireless networks, and smart devices [16], [17], [18]. One of the major challenge associated with healthcare domain is the data access through smart devices. There is an increased chance of malicious attacks or compromised security through these channels. Moreover, when patients are consulting different doctors or various healthcare providers, the sharing of this sensitive information and its privacy and confidentiality is a concern. Hence, while managing the Healthcare data, the multitude of challenges needs to be addressed. The major challenges include but not limited to the authentication of access permission, interoperability of health records, data sharing policies

or rule enforcement on shared content, the secured sharing of health records, and security and privacy concerns about wearables and smart devices for mobile health data capture and processing.

Interoperability in healthcare has been an exciting area of interest for a long time regarding data exchanges between various stakeholders, like insurance providers, researchers, different health care providers, etc. However, this has caused another implication in securing the privacy and identity concerns of the patient. A recent trend in healthcare data management is in a transition phase from organization centric data management towards patient-driven, patient-mediated data management, as patients are true owners of the data. Recent privacy regulations like the General Data Protection Regulation (GDPR) and HIPAA (Health Insurance Portability and Accountability Act of 1996) also demanded a patient mediated or patient-driven data interoperability. In short, the concern mainly revolves around the implementation of a patient-driven data sharing mechanism conserving his privacy, security, and audit-ability of the shared data [9]. However, there exists some ambiguity regarding whether patients should be able to access their entire health data. HIPAA recommends that every individual has the right to access their health data on demand (excluding exceptions, like notes on psychotherapy) [20]. Blockchain is a distributed ledger technology that provides a trusted infrastructure platform for setting up a multi-party business network for mutually distrusting agents. The immutability of records in the shared ledger is achieved through the hashed linking of blocks and the efficient consensus algorithm which makes sure the persisted data is foolproof. All the participants in the network have the same copy of ledger with the same data. The Blockchain mandates the business contracts across the participating parties in the network using the powerful self-executing smart contact that enforces the contractual agreement [21][31].

In this work, we explore the potential of blockchain in leveraging a patient-centric data interoperability architecture for medical records sharing, where data owner decides the granularity of information to be disclosed with intended parties. This work explores the hierarchical architecture using blockchain which captures the traceability of the data access policies set by the data owner, an audit trail of the details of data accessed by third party requesters, incentive agreements with the third-parties for the data sharing, etc. This work also explores the potential of building trusted Artificial intelligent models using e-health data. AI or machine learning models are not trained and tested on the same datasets. They are usually trained on valid authenticated data sets and later used for prediction. This kind of trained data model finds application various domains such as finance industry for volatile market prediction, in the health-care domain for disease identification, anomaly detection in the manufacturing industry and automobile industry, identifying security and privacy violation or intruder detection in the industrial system, etc. For building a good customer base for the AI models, it is necessary that it is trained using trusted datasets and the provenance of these training data to be preserved. Blockchain find a comprehensive solution for capturing the provenance or traceability of the datasets and algorithms used for building the AI models immutably. In this work, we propose a Blockchain based platform for securing e-health data to develop a patient-centric, permission driven data access platform. This work is further extended to build a provenance-based audit trail for the business parties to access the permission given data to run analytics and build trusted AI models in E-health.

A. Contribution of this Article

We have noticed that very few researches have been reported in the Blockchain literature on how to store healthcare data securely and very limited research on patient-driven data sharing and interoperability. To the best of our knowledge, there are no previous works reported that discuss an end-to-end solution on capturing and storing patient information using Blockchain, and also patient-driven data share mechanisms that lead to a provenance-based audit trail for building trusted AI models in Healthcare. In summary, the key contributions made in this article are as follows:

- We propose a framework powered by Blockchain technology that facilitates a patient-driven data sharing mechanism which gives absolute power to the patient to take control over their data.
- We show that a Blockchain powered infrastructure for recording healthcare data including the patient details and medical history lead way for enabling a trusted Artificial Intelligence ecosystem that enhance current centralized AI models.
- We also extend this work for building a provenance-based audit trail for businesses to access the permission on patient-centric data to build trusted AI models and run analytics without disclosing personally identifiable information.

B. Organization of this Article

This article is organized as follows: Section II discusses the preliminaries and background for this study, primarily Blockchain and Electronic Health (e-Health). Section III describes the state-of-the-art of data privacy and security using the Blockchain and also some related works that throw lights on decentralized artificial intelligence which is an emerging area in computing. In section IV, we introduce our proposed frameworks for healthcare data capture, storage, and sharing mechanisms and section V discusses how the proposed frameworks enable a trusted artificial intelligence ecosystem. Section VI concludes the article and our outlook is explained.

II. Preliminaries and Assumptions

A. Blockchain Technology

Blockchain is a distributed ledger technology that provides a secured, immutable, tamper-proofed, distributed data store. Blockchain provides a trustless platform that offers a transparent infrastructure network where a group of non-trusting parties can collaborate and do business without any third party authorization. Instead of centralized record keeping of transaction storage, blockchain offers a decentralized data-store, where every peer in the network keeps a copy of replicated data. The replication, sharing, and synchronization of the data across the peers are managed by the network through some consensus mechanism. The distributed ledger is built on top of a collection of communication protocols that enables decentralized administration of transactions that persisted across geographically separated multiple computing devices. The first application of blockchain is proposed by Satoshi Nakamoto [13], to provide a distributed, transparent technology platform for the much-celebrated cryptocurrency, Bitcoin. The immutability of the data is managed by the consensus mechanism and efficient hashed block linking mechanism. Consensus algorithm makes sure all participating peers have the same state of the network and multiple rounds of endorsements are required for any piece of data to go into a block. A diagram depicting the high level structure of Blockchain is shown in Fig. 1.

The combination of cryptographic hashes, structured hash linking design , multi-party consensus mechanism, and replicated data storage across the nodes provide a trust-full platform by blockchain. A block in Blockchain encapsulates an ordered transaction set, the hash of the previous block, and optionally the nonce of the validating peer [27]. A nonce is nothing but a uniquely identifying piece of information in block creation. In blockchains like Bitcoin or Ethereum, Nonce is the solution to the complex cryptographic puzzle solved by the miner. Every committed block of transactions holds the hash of previous blocks as input. This brings in the immutability of the data records. If

any member in the network tampers their copy of shared data, the hash generated from the malicious node will be different from other peers in the network. This causes the falsified node to fail the consensus validation. Unless they are not synced with the rest of the network, they cease to exist in the network. Blockchains are broadly classified into three categories namely public Blockchains, private Blockchains, and consortium Blockchains.

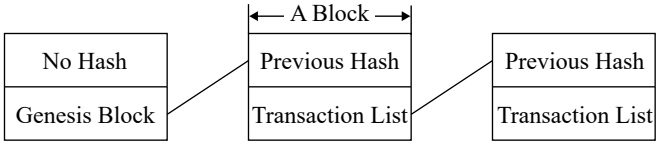


Fig 1. Structural organization of Blockchain.

Public Blockchains are open access blockchain implementations. Membership services are not required for enrolling new users. Any user can be on the blockchain. It is open, which means that any node can access the contents(transactions) of the chain without any restrictions. Any node can take part in the consensus process. The consensus process determines which transactions get added to the Blockchain and maintains the current state of the system. Public blockchains are decentralized meaning no single entity has control over the network. Private Blockchains are privately managed by an individual or organization and only registered members could be part of it. In a private Blockchain, transaction validations are managed by consensus mechanisms or the policy or rules that defines the endorsers or validators. For example, Ethereum Private Network. On the other hand, Consortium blockchains are blockchains managed by a group of institutions who owns a pre-selected number of nodes. These blockchains are partially decentralized [29],e.g.: Corda.

Another classification is permissioned Blockchain and permissionless Blockchain. In permissioned Blockchain, the role of the user can be defined and only the defined user according to the policy has access to the Transaction. For example, Bitcoin and Ethereum are permissionless Blockchains where anyone can be part of the network, the only mandate is he should follow the Proof-of-Work consensus mechanism. However for an enterprise application, to manage the privacy and confidentiality of the data and transaction, industries prefer customizable, access permissions, resource effective Blockchain implementation. This led to the development of permissioned Blockchain [25], [26], [27], [28], [29], [30], e.g.: Hyperledger Fabric, Hyperledger Sawtooth, etc.

B. e-Health and Blockchain

E-health is a term generally used to refer to information and communication technology (ICT) in the healthcare sector which has been evolving since 2000. It has the potential to provide innovative solutions in the medical field encompassing domains like medical informatics, patient healthcare, health care providers, health care professionals, pharmacy, etc. referring to health services, health information, etc. through the Internet and related technologies [33]. As information or data availability becomes seamless through extended Application Programming Interfaces (APIs) and smart devices, patients have a handle to access their medical records and treatment information. Moreover, the data is still getting generated and residing in organizational silos, opening electronic access for the patient to view and manage their data and share it with appropriate stakeholders on demand. The recent trend in healthcare data management is transitioning from organization driven data management towards patient-driven data interoperability to be in-line with GDPR compliance and HIPAA compliance, which demands patients to be the true data owners and not the institutions that generate or create the wealth of data. However, moving towards patient mediated data handling would need to address challenges like implementing new security and privacy protocols for data sharing, authentication, interoperability, confidentiality requirements, electronically verifiable consents, and policy enforcement and governance for data management [9], [1].

C. Securing e-health System and Efficient Data Storage and Sharing

Storage and handling of medical data require a clear definition of access control, authentication, and immutability of data records for ensuring integrity, confidentiality, and accessibility of the health records. Health records encompass both wearable sensor data as well as patient treatment records from healthcare providers. As medical record keeping needs privacy and confidentiality due to data sensitivity, role-based authorization should be implemented for data upload as well as data access. Auditing mechanisms should be enforced to record and monitor the queries and data accesses to be monitored.

Blockchain provides an efficient framework for setting up a private/public network with a secured data access control platform for multi-party business. The potential of Blockchain to create a secured, immutable data sharing platform has been explored previously. Yue et. al. described a methodology of employing a private blockchain for medical record booking and implementing Healthcare Data Gateway (HDG) to enable patients to manage their health records [35]. Ivan [36] proposed a public Blockchain implementation of Personal health record keeping, where medical records are encrypted and stored publicly in a distributed fashion. MedChain [37] employs a permissioned distributed ledger network, where the healthcare stakeholders such as patients, healthcare providers/professionals, pharmacies, etc., could be used to facilitate medication-specific data sharing between agreed entities.

However blockchain technology, has its own share of implications in putting the data records on blockchain. Hence there are few attempts of leveraging blockchain as an enabler for trusted infrastructure in healthcare. They did not assume blockchain as a data layer but, but used smart contracts and blockchain for facilitating management or governance of that data. Zyskind et al. [38] have proposed a blockchain based decentralized access control for managing and handling of encrypted off-chain data. FHIRChain [39], [40] is a healthcare usecase, which is a smart-contract managed system for health data exchange, where the actual clinical data is not residing in the blockchain. But the meta-data of the the clinical records are encrypted and maintained in blockchain as a pointer to actual records in the cloud. Azaria et al. [42] proposed MedRec, which is a prototype built on permissioned blockchain network to enable and manage data sharing and authentication process. MedRec has a novel incentivisation method built around to access the anonymized medical data.

III. Data Privacy and Security using Blockchain: Current State

A distributed ledger management infrastructure implemented for establishing trust in the popular crypto-currency - Bitcoin is now disrupting how businesses function. Trust, being the primary asset of any multi-party involved entities, is achieved by Blockchain by creating a hash-chained record of transactions in an immutable fashion. This is a disruptive technology with many interesting and useful real-world applications that require a trusted data management ecosystem. A cluster of computers (nodes) that decentralize the data management capabilities of an organization, the Blockchain ecosystem, now have the capabilities and potential to move the notion of trust from institutions to mathematically provable infrastructure. This capability exhibited many inherent features that build a trusted platform that captures audit-free transaction data which is essential for any business. This attracted many

organizations to implement this technology to enhance their business models and as a result, many variants of Blockchain have emerged with varying levels of success and usability. This section outlines some of the recent advances in using Blockchain technology for better data privacy and security enablement in businesses and other domains.

Very recently, a Blockchain based infrastructure for air traffic management security, authentication, and privacy has been introduced by Ronald J. Reisman [2] which is an open source, permissioned Blockchain based framework. This design enables aircraft privacy and anonymity while providing an efficient and secure mechanism for communicating with air traffic operations support and other stakeholders or entities. This framework provides better security and privacy by enabling higher bandwidth communication channels for private information and deploying smart contracts [2]. The connected and decentralized IoT devices are growing in an exponential fashion which exposes high risk for businesses in terms of privacy and security. The adoption rates for such smart devices are highly dependent on the ability of vendors to provide sufficient sensor data integrity while preserving the privacy of users. A new method for privacy-preserving data certification in IoT networks by leveraging Blockchain technology for protecting sensor data has been introduced recently [3]. The system proposed a blockchain based sensor data protection framework for certifying the data generated and propagated by sensors in an IoT network. The authors claim that the proposed design ensures the tamper-resistant gathering, processing, and exchange of IoT (Internet of Things) sensor data in a privacy-preserving, scalable, and efficient manner [3].

A privacy-friendly platform for healthcare data in cloud-based on Blockchain environment [4] is introduced in the literature that presented a patient-centric healthcare data management system powered by Blockchain technology for attaining privacy. This work used cryptographic functions for encrypting patient's data for ensuring pseudo-anonymity [4]. A novel framework for addressing security and privacy issues in healthcare using modified Blockchain models which are apt for IoT devices was introduced by Ashutosh Dhar Dwivedi et. al. [5]. The additional security and privacy features of their modified Blockchain model are based on basic cryptographic properties that make IoT based healthcare networks more secure and anonymous [5]. A Blockchain based learning healthcare system that could foster the willingness of a patient to contribute to research and learning by providing sufficient control over health data is proposed by Marielle S. Gross and Robert C. Miller [6]. This system also enables mechanisms for meta-analysis without even exposing individual details of the patient and this will allow protecting the privacy of patients [6].

Another system for medical data management on Blockchain for preserving privacy, that uses Hyperledger Fabric to store encrypted data has been proposed by Haibo Tian, Jiejie He and Yong Ding [7]. The authors proposed the idea of a shared key that could be reconstructed by legitimate parties before starting the treatment of patients. The approach uses sibling interactable function families to establish the shared key for data encryption [7]. Design of a storage scheme for managing and sharing medical records is proposed by Yi Chen et. al. The storage and sharing schemes employed by the proposed system do not depend on any third-party and there is no single entity that has absolute power to make the process affected [8].

William J. Gordon and Christian Catalini looked at how Blockchain technology can facilitate patient-driven interoperability by using features such as digital access rules, data aggregation, data liquidity, patient identity, and data immutability [9]. The authors concluded that patient-driven interoperability is a trendsetter in the healthcare industry having many challenges, Blockchain may facilitate the transition from a healthcare provider centric ecosystem to a patient-centric ecosystem [9]. A new framework called ModelChain was introduced to adapt Blockchain technology for privacy-preserving machine learning systems [10]. In this system, the machine learning model parameter estimation is done without disclosing any patient health information and the proposed framework applies Blockchain technology to solve the privacy-preserving predictive modeling task for healthcare that will facilitate and increase the potential of interoperability between various stakeholders [10]. A new data sharing solution built on Blockchain that addresses two prevalent issues with the healthcare domain such as protecting sensitive health information and deployment and installation of Blockchain software across diverse hospital networks has been proposed [11]. The newly proposed innovative architecture addressed critical data security, deployment, and installation challenges and provides the healthcare community with an approach to connect diverse and heterogeneous providers while protecting sensitive healthcare data [11].

A. Trend Towards Trusted and Decentralized Intelligence

Advances in computing and communication technologies and heavy investments in exponential technologies such as Artificial Intelligence caused hundreds of millions of devices to run AI / ML models. But these devices and AI models are built and run by some centralized mega-corporations that makes the entry difficult for other small entities into this ecosystem. The organizations who collect and curate data may not be running the AI models most of the time and the end users will be mostly another set of people. This situation demands Gabriel Axel Montes and Ben Goertzele in their work on Distributed, decentralized, and democratized artificial intelligence [14] claims that decentralizing the AIwill generate more equitable development of Artificial Intelligence and Artificial General Intelligence. The authors also state that the progress towards decentralized AI will also create the infrastructure for coordinated action between AIs that will significantly facilitate the evolution of AI into true Artificial General Intelligence that is both highly capable and beneficial for humanity [14]. A recent study on the decentralized marketplace data using Blockchain technology which aims to provide confidence among multiple stakeholders in the system. This specific study on the supply chain domain ensures four different levels of data management such as data provision, data delivery, rights management, and producer internal sources [33].

In this work, we propose a Blockchain powered framework for securely storing patient sensitive information and other associated healthcare data such as treatment history, by enabling a patient-driven data sharing mechanism. The core idea of the proposed work is to make the patient the true owner of data and thereby giving absolute power to decide which data needs to be shared with the healthcare providers. The audit trail of the complete transaction details will be recorded in a tamper-proof digital ledger which is an auditable record for reference. The digital identity management of the proposed framework will generate a unique identity for a patient to which the treatment history and other medical records can be appended and this makes a moving electronic health record management ecosystem which is patient-managed. In addition to this, we propose how a Blockchain based data management framework that employs audit trail features can enable the notion of trust on the data captured among multiple entities to make use of a trusted but decentralized AI model.

IV. Blockchain for Securing Healthcare Data

Recent researches on data management in healthcare shows a clear transition from an organization (healthcare provider) centered data management to patient-driven data management mechanisms. Even though the idea of transforming patients as the true owners of their data is not new, the very recent regulations such as the European Union General Data Protection Regulation (GDPR) fueled this transformation. According to European Union, the GDPR (https://eugdpr.org/), is the most important and revolutionary change in data privacy regulation in the last twenty years which will reshape how

data is handed across every sector - be it healthcare to banking and beyond. While many other sectors will change how they manage data, everybody is eyeing the healthcare domain and is very keen to analyze how the regulations mentioned above are going to be implemented in healthcare. We can see that regulations in healthcare such as the 2009 Health Information Technology for Economic and Clinical Health Act (HITECH) introduced in the United States has increased the digitization of Electronic Health Records (EHR), the sharing of such records among different hospitals and other healthcare providers are lagging much behind. This is due to the operational, technical and most importantly the privacy-related issues prevalent specifically in the healthcare domain [9].

Interoperability in healthcare is often interpreted as the ability to exchange medical records electronically between hospitals managed by the same business unit or between multiple individual hospitals, to some extent. Here, the data management, protection, and exchange is the responsibility of the business units such as hospitals and the patient normally does not come into the picture. In such a scenario, the patients have no clue where their data is being stored or with whom their data has been shared. Recently, there is a paradigm shift from the traditional business-centric approaches to a patient-driven or patient-centric interoperability where the patient becomes the absolute owner of their data. However, this shift introduces an array of new challenges in terms of technology, security, privacy and governance and still many of the issues related to these aspects have not been solved for even business-centered interoperability use-cases. To accelerate the facelifting of patient-centered data management, we need new technology supported initiatives and systems that not only solve the issues associated with the traditional interoperability but also pave ways for patient-centered interoperability easily.

Blockchain, the technology infrastructure behind Bitcoin[13], is a novel technology that has built-in features such as tamper-proof, audit-ability, decentralization, cryptographic signatures and hashing, etc. makes it appropriate for storing data securely which leads to improved interoperability[12][13]. Blockchain technology has the potential to facilitate and accelerate the transition from an institution-centered data management ecosystem to patient-centered or patient-controlled data management by incorporating data aggregation, data access rules, immutability, and patient identity management[9]. Identifying and incorporating the benefits provided by Blockchain technology in facilitating privacy preserving and patient-centered data management,

TABLE I. Features of Blockchain and Application in Patient Driven Data Management and Data Sharing

feature	APPLICATION
Immutability	Patient information, treatment information and other healthcare records can be securely shared among multiple providers lowering the risk of attacks and loss and thus avoiding the need for audit-ability and verifiability.
Identity Management	Every patient can have a unique digital identifier which can be stored efficiently on Identity Management Blockchain and all the EHR and other treatment information can be mapped against this identity without the worry of being tampered
Decentralization	Since there is no notion of centrality, the clinical Decentralization and other treatment data can be accessed among multiple hospitals and near realtime records can be accessed.
Access Rules	Patients can give consent for accessing their Access Rules medical records through self-executing codes on blockchain called “smart contracts” and which in turn can keep audit trail of data access transactions

we have attempted this problem in detail and proposed new frameworks for patient data management and sharing. We have also extended these frameworks for building a provenance based Blockchain model for an audit trail mechanism that leads towards a trusted artificial intelligence ecosystem for healthcare. A list of blockchain features and their application in patient-driven data sharing mechanisms are given in Table I.

A. Blockchain based Data Privacy and Security Systems

Here, we propose a Blockchain based digital identity and consent management mechanism for healthcare which is shown in Fig.2.

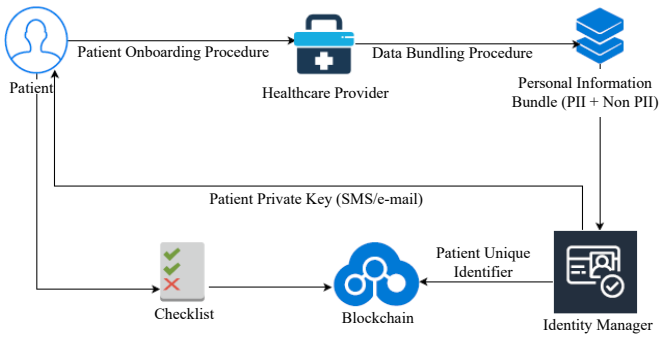


Fig. 2. Proposed Patient Registration / On-boarding Process.

- **Patient Registration / On-boarding Procedure:** The healthcare provider (HCP) collects the complete information including personally identifiable information (PII) and non-personally identifiable information (non-PII) of the patient through the patient on-boarding procedure. The HCP then segregates PII and Non-PII elements from the data and creates a data bundle (PII + Non-PII) and passes this information into a Blockchain based identity manager. If the hospital or healthcare provider already has a database based EHR management system, then the Non- PII data can be pushed to the same system. The personally identifiable information will be captured in an offline database and the hash of this record will be stored in the Blockchain. In order to link the offline record with the Blockchain transactions, we employ a foreign key generated using a hash function. This allows the patient to update the personally identifiable information seamlessly. However, we only describe a complete blockchain based system assuming that the hospital does not already have a technology infrastructure for dealing with patient data and need a complete tamper-proof mechanism to deal with patient information and other electronic health records. The identity manager (IM) module of the proposed framework generates a unique identifier for the on-boarded patient and record the same in the Blockchain. Simultaneously, the IM will generate a private key for the patient and communicate the same through an SMS or e-mail to the registered mobile phone number or e-mail address of the patient. During onboarding or registration, the healthcare provider gives a checklist of data that needs to be captured and/or shared and the patient needs to check/un-check the list of items according to their willingness to get the data captured by the provider. On submitting the final checklist, the same will be updated in the Blockchain against the private key of the patient. Finally, a patient will have a unique identifier and an associated checklist of items related to their data sharing consents. Such a system would facilitate a patient-centered and patient-driven data sharing ecosystem where the patient has absolute power on their data and only with his consent, the data can be viewed or shared among various healthcare providers. Revenue models can be built around data sharing platforms. Based on the permission agreement

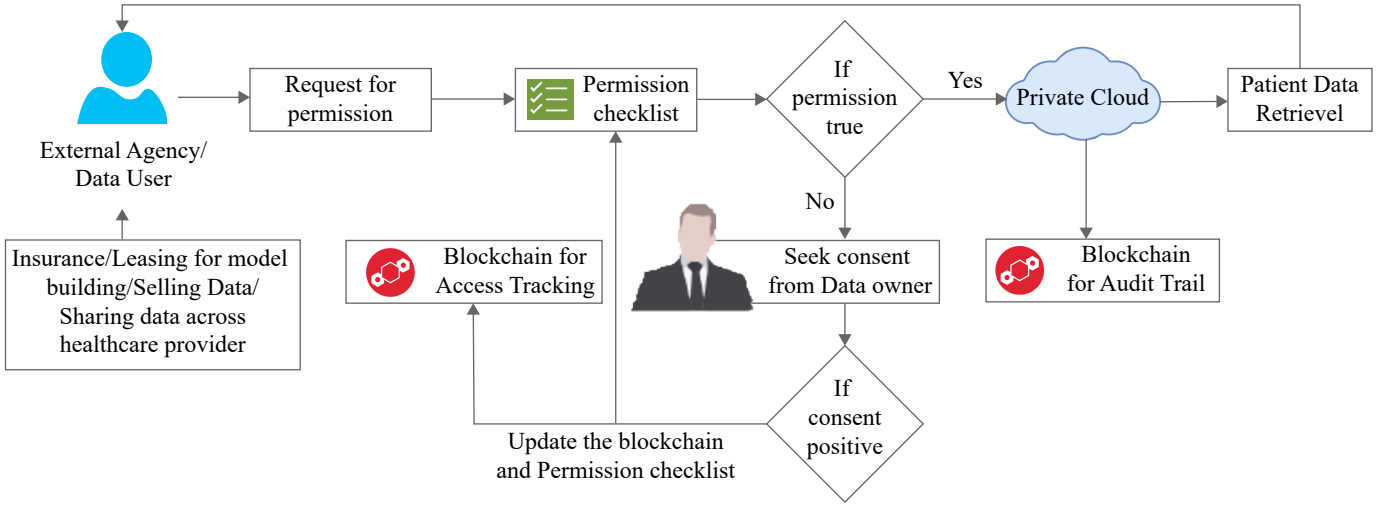


Fig. 3. Patient Data Request Process.

data could be shared across various stakeholders under stipulated pre-published pricing policies. Blockchain offers excellent platform for immutable record keeping of the rule based condition of data sharing policy and tamper-proof distributed compute-storage, which provides a secure environment for executing multi-party business. Self-executing smart contracts mandates the policy enforcement and the audit trial logging mechanism in blockchain offers the traceability of all executed transactions and smart contracts.

- **Data Request Procedure:** In this proposed system, patient health records have been captured and stored in the Blockchain with an appropriate checklist of permissions. When a healthcare provider or other users request for this recorded data, they need consent from the owner of the data. To enable this, a request for granting the permission will be submitted to the network with a list of required data items. The requested data records will be cross-checked with the patient published checklist. Consented data items for which the patient has given consent will be shared with the requester. Simultaneously, the audit trail of the data access along with complete request information will be stored in a Blockchain exclusively for audit trail. The data owner will also be alerted regarding data access. In case of data access permission revocation, a change in the permission will be recorded in the audit trail Blockchain as a new transaction and an intimation will be sent to the data requester. The proposed data request for healthcare providers is shown in Fig.4.

In case the data requester/user is requesting data items which have no access rights/consent given by the patient, then this request will be sent to the patient asking for his consent. It is the sole discretion of the patient to grant access to the data items. If the patient gives a new consent for the requested item, then the same will be first updated in the permission checklist and also on the Blockchain for access tracking. In this way, the patient can grant or deny access to data items that are owned by him.

B. Towards a Trusted Decentralized Artificial Intelligence Ecosystem using Blockchain

Heavy investments in Artificial Intelligence and Machine Learning caused trained ML models to become more accessible by various stakeholders which in turn benefits the end users directly and indirectly. In such a decentralized intelligent ecosystem, the various participants need to trust the model which requires provenance information on how the data has been captured and trained. According to Sarpatwaret. al. [41] the provenance data model for an AI/ML model includes the following information:

- Complete details of the data used to build the AI model.
- Details of the model pipeline.
- Information on the training process.
- Details of the updates to the trained model.
- Testing details of the trained model.

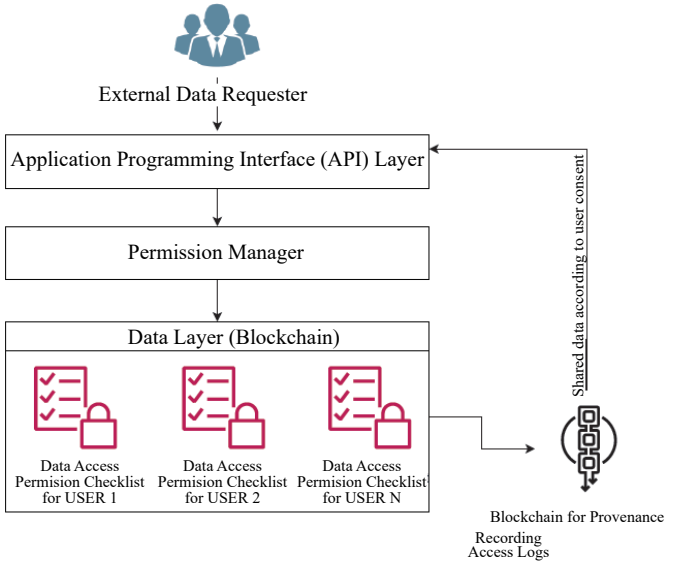


Fig. 4. Data Request Process for Healthcare Providers.

This information needs to be stored in a tamper-proof, decentralized and immutable manner to harness the trust of the decentralized AI model. Blockchain is an ideal solution to tackle this situation which can capture the complete information about the trained model right from the data collection to the model validation to testing. This facilitates and creates a transparent and trusted decentralized artificial intelligence ecosystem where various participants share the credible data or trained model. Finally, there won't be a trusted centralized organization that manages the AI model generation; rather there will be decentralized and trusted participants that can contribute to the overall training and evaluation of AI models. The healthcare domain is not an exception, thus we need a Blockchain based provenance mechanism for capturing patient data right from onboarding (with appropriate

consent from the data owner) which is powered by Blockchain. We propose a framework for provenance (Refer Fig. 3) that aims to provide or give access to captured healthcare data for external vendors to build Artificial Intelligence / Machine Learning models or run analytics. This is done by exposing REST APIs with appropriate authentication and permission management procedures. The data layer stores and manages data access permission checklist for every individual patient onboard in the system. On accessing data after getting sufficient consent on the data items requested, complete access logs will be recorded in the provenance Blockchain which makes the data credible for other vendors to build AI/ML models which will be beneficial for many number of services and applications.

C. Experimental Setup

Here, the implementation details of our proposed frameworks are discussed. We have created an experimental testbed using Hyperledger Sawtooth which is an enterprise Blockchain platform for creating and managing permissioned ledgers. We have used the native installation using Ubuntu 16.04 with Sawtooth version 1.0 which is available at https://github.com/hyperledger/sawtooth-core/ for simulating a private Blockchain network locally. Since we need to implement activities of different stakeholders such as healthcare providers, patients, and regulators, a three node network with three validators have been configured in the network. Also, Sawtooth validators, REST APIs and transaction processors are created and initiated in the already created network. We have used Python SDK for Sawtooth available at https://github.com/hyperledger/sawtooth-sdk-python. For comparing the operational advantages of a database based system w.r.t a blockchain driven health- care data managed system, we setup a MySQL database, where the DB provide the single, centralized instance of data source. However this DB based system is not free from data-manipulation or data forging attempts by malicious agents.

In a blockchain driven system, data could be either stored off-chain and the meta-data of the stored information could be immutably secured in blockchain. However in this work we loaded the data directly into blockchain and access permissions to the data are managed by smart contracts. The subtle advantage of the blockchain driven system over database only based system is the immutability of the data records in blockchain, be it permission settings permitted by the patient, data access audit trial, data logging into the system, etc. Revenue models built around it would need a transparent eco-system

D. Dataset Description

The main challenge faced by researchers in the healthcare domain is the availability of public datasets. Since healthcare records are very sensitive, they are mostly not available publicly. However, there are very few anonymized datasets accessible online. The proposed work requires data records with patient identifiers such as name, date of birth and other contact details, the anonymized datasets will not suit our needs. For this work, we have used a hybrid method for collecting anonymized data records and appended dummy data records with personally identifiable information. The MIMIC-III dataset is a freely available healthcare dataset containing deidentified health-related data associated with over 40000 patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012 [32][33]. For the current experiment, we For the current experiment, we have only used 1000 records due to scalability and other data management constraints.

E. Implementation Details

In our proposed framework, the patients need not run a network node and they are treated as clients. Infrastructure will be provide REST APIs for accessing and updating their personal information, data access permissions and other consents. We have implemented a web application that captures complete data of patients including PII and non-PII while on-boarding at a health-care provider. For this experiment, we have on-boarded 1000 patient records taken from MIMIC-III dataset (after appending identifiable information) and has recorded the same in the Blockchain (Refer Fig. 1) along with default consents for all records. In order to compare the operation of a simple database system with respect to a Blockchain mediated system, we have implemented a standalone MySQL database (DB) for record keeping. MySQL database holds patients personal details in patient-personal table and treatment details in imaging table, diagnosis table and pharmacy table. Access permissions are stored in *usr-access-request* and *usr-access-perm* tables. Data logging regarding the permissions set by the user and details of the data access by the user are persisted in audit data table.

Our data request framework (Fig. 3) requires two separate Blockchains: One will store patient information and data access consents and another one for capturing the audit trail of the queries/ requests made by any requester. We tried accessing information of different patients and found that the data requester (client) can only access those data items from the patient record for which the patient has given access to. In addition to this, we have also tested whether access logs are added to the Blockchain for audit trail. We later updated the default data access permission of patients and repeated the data request process. This is to ensure that the changes are updated in the system while updates are made into the system through the checklist API provided.

Finally, we have set-up provenance Blockchain using Hyperledger Sawtooth for implementing our framework for external data request by third parties. Whenever a data request has been initiated by external data requester, the permission manager checks and verify whether the requester has sufficient privileges to use the REST APIs for data access. To check this functionality, we have added another client to the provenance Blockchain network and given access to the APIs. When a data access request is received, the individual data access permission (patient consent checklist) is referenced and on granting data records, complete access logs including the requester details and granular level data details will be recorded in the Provenance Blockchain. This ledger keeps track of all data access information which will act as a credible source of data on which decentralized artificial intelligence or machine learning models can be built and share.

In-order to compare the operation of a blockchain based system with other legacy system, we assume, a malicious actor from the healthcare provider is trying to tamper with the data. This is achieved by manipulating the access permissions set by the patient and attempts to alter the pricing strategy to benefit the organisation. However as blockchain maintains immutable and tamper proof logging and record keeping of the transactions in distributed manner, any malicious attempts in local data manipulation, would be caught during consensus evaluation. Also, every attempt to login into the system is also captured in the audit module, details of all unauthorized access to the patient data can be analyzed in future.

V. Discussions and Lessons Learnt

The proposed approach was implemented in a controlled environment with a restricted number of user profiles. Blockchain technology for enterprise applications are still in an infancy stage. We faced issues on integrating various aspects of the proposed system in terms of scalability, usability, accessibility and interoperability. The key challenges identified during our pilot implementation of the proposed method are outlined below:

- Interoperability between cloud, legacy system, mobile application and Blockchain.

- Organizational policies and restrictions on storing data over cloud.
- The inhibition from patients to store and acquire data by the cloud services and the associated cost for the same.
- Absence of data protection regulations and compliances from the regulatory bodies in healthcare.

Comparison Note: Current centralized systems use dedicated database servers deployed in organization premises or cloud-based services which brings the notional of centrality to a greater extent. Maintaining a single version of the truth and deriving actionable insights from data accumulated at local repositories is always a complex task. The extraction, transformation, and loading (ETL) process of data from local repositories are complex and thus a major hindrance to the quick decision-making process. On the other hand, a Blockchain always has a single version of the truth that is maintained in a decentralized manner. As every node in the network has the same copy of ledger which is audit-free, the near real-time analysis can be performed. While applications that use database backend show better performance and scalability, the major bottleneck of Blockchain is found to be scalability, interoperability and the delay in committing transactions. The technology is still in an infancy stage and requires a lot of improvements in terms of scalability and interoperability that may enable higher transaction throughput and better-managed services.

VI. Conclusions and Future Work

This work proposes a platform that enables patients to be the true owner of their data. The frameworks support secure, immutable, auditable and traceable platform powered by Blockchain that enables patients to manage their healthcare data. We show that our proposed frameworks can take away the notion of trust from centralized organizational infrastructure to decentralized mechanisms. The system also caters to the requirements for building a trusted AI ecosystem with provenance. The Proof-of-Concept (PoC) implementation of our proposed frameworks in a Permissioned Blockchain shows that there is significant potential to explore further. This PoC with limited data management capabilities needs to be scaled-up to incorporate new stakeholders and other entities. Patient mobility and easiness of managing permissions are issues that need to be addressed and the authors will be working on these dimensions in the future.

Acknowledgment

This work is supported by the Back to Lab Programme research fellowship (Ref: Order No.1281/2016/KSCSTE) from Women Scientists Division (WSD), Kerala State Council for Science, Technology and Environment (KSCSTE).The authors would like to thank all the researchers, engineers and other staff members of Kerala Blockchain Academy and Data Engineering Lab at Indian Institute of Information Technology and Management-Kerala, for their valuable comments that significantly improved the quality of this paper. The authors also acknowledge the staff members for providing infrastructure facilities for conducting an experiment of this scale.

References

[1] T. Mcghin, K.-K. R. Choo, C. Z. Liu, and D. He, "Blockchain in healthcare applications: Research challenges and opportunities," *Journal of Network and Computer Applications*, vol. 135, pp. 62–75, 2019.

[2] R. Reisman, "Blockchain Serverless Public/Private Key Infrastructure for ADS-B Security, Authentication, and Privacy," *AIAA Scitech 2019 Forum*, 2019.

[3] M. Chanson, A.Bilgeri, D., Fleisch, E., and Wortmann, F. "Privacy-Preserving Data Certification in the Internet of Things: Leveraging Blockchain Technology to Protect Sensor Data," *Journal of the Association for Information Systems,* 2019.

[4] A. A. Omar, M. Z. A. Bhuiyan, A. Basu, S. Kiyomoto, and M. S. Rahman, "Privacy-friendly platform for healthcare data in cloud based on blockchain environment," *Future Generation Computer Systems*, vol. 95, pp. 511–521, 2019.

[5] A. Dwivedi, G. Srivastava, S. Dhar, and R. Singh, "A Decentralized Privacy-Preserving Healthcare Blockchain for IoT," *Sensors*, vol. 19, no. 2, p. 326, 2019.

[6] M. S. Gross and R. C. Miller, "Ethical Implementation of the Learning Healthcare System with Blockchain Technology," *Blockchain in Healthcare Today*, vol. 2, 2019.

[7] H. Tian, J. He, and Y. Ding, "Medical Data Management on Blockchain with Privacy," *Journal of medical systems*, vol. 43, no. 2, article 26, 2019.

[8] Y. Chen, S. Ding, Z. Xu, H. Zheng, and S. Yang, "Blockchain-Based Medical Records Secure Storage and Medical Service Framework," *Journal of Medical Systems*, vol. 43, no. 1, 2018.

[9] W. J. Gordon and C. Catalini, "Blockchain Technology for Healthcare: Facilitating the Transition to Patient-Driven Interoperability," *Computational and Structural Biotechnology Journal*, vol. 16, pp. 224–230, 2018.

[10] T. T. Kuo, and L. Ohno-Machado (2018), "Modelchain: Decentralized privacy-preserving healthcare predictive modeling framework on private blockchain networks," arXiv preprint arXiv:1802.01746.

[11] M. A. Cyran, "Blockchain as a Foundation for Sharing Healthcare Data," *Blockchain in Healthcare Today*, 2018.

[12] C. Catalini and J. Gans, "Some Simple Economics of the Blockchain," 2016.

[13] S. Nakamoto"Bitcoin: A peer-to-peer electronic cash system.", 2008

[14] G. A. Montes and B. Goertzel, "Distributed, decentralized, and democratized artificial intelligence," *Technological Forecasting and Social Change*, vol. 141, pp. 354–358, 2019.

[15] What Is Patient Engagement? | Evariant: The Leading Healthcare CRM Solution. [online] Available at: <https://www.evariant.com/faq/why-is-healthcare-data-management-important> [Accessed 23 June 2020].

[16] H. Löhr, A.-R. Sadeghi, and M. Winandy, "Securing the e-health cloud," *Proceedings of the ACM international conference on Health informatics - IHI 10*, 2010.

[17] D. Kotz, C. A. Gunter, S. Kumar, and J. P. Weiner, "Privacy and Security in Mobile Health: A Research Agenda," *Computer*, vol. 49, no. 6, pp. 22–30, 2016.

[18] M.A.Sahi, H.Abbas, K.Saleem, X.Yang, A.Derhab, M.A.Orgun, W. Iqbal, I. Rashid, and A. Yaseen, " Privacy preservation in e-healthcare environments: State of the art and future directions," *IEEE Access*, vol. 6, pp.464-478. 2017.

[19] M. Panner, "Blockchain In Healthcare: How It Could Make Digital Healthcare Safer And More Innovative," 2019, <https://www.forbes.com/sites/forbestechcouncil/2019/06/18/blockchain-in-healthcare-how-it-could-make-digital-healthcare-safer-and-more-innovative/#2ff4bacf3e5a> [Accessed 23 June 2020].

[20] W. J. Gordon and C. Catalini, "Blockchain Technology for Healthcare: Facilitating the Transition to Patient-Driven Interoperability," *Computational and Structural Biotechnology Journal*, vol. 16, pp. 224–230, 2018.

[21] W.J. Gordon, A. Wright and A. Landman. Blockchain Technology in Health Care: Decoding the hype. NEJM Catal2017 https://catalyst.nejm.org/decoding-blockchaintechnology-health/, Accessed date: 28 March 2018.

[22] A. J. Holmgren, V. Patel, and J. Adler-Milstein, "Progress In Interoperability: Measuring US Hospitals' Engagement In Sharing Patient Data," *Health Affairs*, vol. 36, no. 10, pp. 1820–1827, 2017.

[23] Q. Nasir, I.A. Qasse, M. Abu Talib, and A.B Nassif, "Performance analysis of hyperledger fabric platforms," *Security and Communication Networks*, 2018.

[24] A. Brando,H. So Mamede,and R. Gonalves (2019, April), "Trusted Dataset Marketplace," In *World Conference on Information Systems and Technologies* (pp. 515-527). Springer, Cham.

[25] C. Cachin (2016, July), "Architecture of the hyperledger blockchain fabric," In *Workshop on distributed cryptocurrencies and consensus ledgers* (Vol. 310, No. 4).

[26] E. Androulaki, A. Barger, V. Bortnikov, C. Cachin, K. Christidis, A. D. Caro, D. Enyeart, C. Ferris, G. Laventman, Y. Manevich, S. Muralidharan, C. Murthy, B. Nguyen, M. Sethi, G. Singh, K. Smith, A. Sorniotti, C. Stathakopoulou, M. Vukolić, S. W. Cocco, and J. Yellick, "Hyperledger fabric," *Proceedings of the Thirteenth EuroSys Conference*, 2018.

[27] V. Dhillon, D. Metcalf, and M. Hooper, "The Hyperledger Project," *Blockchain Enabled Applications*, pp. 139–149, 2017.M. Swan (2015). *Blockchain: Blueprint for a new economy*. " O'Reilly Media, Inc."

[28] Gorenflo, Christian, Stephen Lee, Lukasz Golab, and Srinivasan Keshav. "Fastfabric: Scaling hyperledger fabric to 20,000 transactions per second." In 2019 *IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, pp. 455-463. IEEE, 2019.

[29] Benhamouda, F., Halevi, S. and Halevi, T., 2019. Supporting private data on hyperledger fabric with secure multiparty computation. *IBM Journal of Research and Development*, *63*(2/3), pp. 3:1-3:8.

[30] K.Olson, M.Bowman, J. Mitchell, S. Amundson, D. Middleton, and C. Montgomery, "Sawtooth: An Introduction. The Linux Foundation," 2018.

[31] I. C. Lin and T. C. Liao, "A survey of blockchain security issues and challenges," *IJ Network Security*, vol. 19, no. 5, 653-659. 2017.

[32] Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. "MIMIC-III, a freely accessible critical care database," *Scientific Data*, vol. 3, article 160035, 2016.

[33] Johnson, A., Pollard, T., & Mark, R. (2019). MIMIC-III Clinical Database Demo (version 1.4). *PhysioNet*. https://doi.org/10.13026/C2HM2Q.

[34] H. Oh, A. Jadad, C. Rizo, M. Enkin, J. Powell, and C. Pagliari, "What Is eHealth (3): A Systematic Review of Published Definitions," *Journal of Medical Internet Research*, vol. 7, no. 1, 2005.

[35] X. Yue, H. Wang, D. Jin, M. Li, and W. Jiang, "Healthcare Data Gateways: Found Healthcare Intelligence on Blockchain with Novel Privacy Risk Control," Journal of Medical Systems, vol. 40, no. 10, 2016.

[36] D. Ivan. "Moving toward a blockchain-based method for the secure storage of patient records", *NIST/ONC*; 2016.

[37] WJ. Gordon, A. Landman, "Secure, decentralized, interoperable medication reconciliation using the Blockchain",*NIST/ONC*, 2016.

[38] G. Zyskind, O. Nathan, and A. sandy Pentland, "Decentralizing Privacy: Using Blockchain to Protect Personal Data," *2015 IEEE Security and Privacy Workshops*, 2015.

[39] "Health Level Seven International," *Health Level Seven International - Homepage*. [Online]. Available: http://www.hl7.org/. [Accessed: 22-Jun-2020].

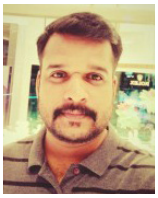
[40] P. Zhang, J. White, D. C. Schmidt, G. Lenz, and S. T. Rosenbloom, "FHIRChain: Applying Blockchain to Securely and Scalably Share Clinical Data," *Computational and Structural Biotechnology Journal*, vol. 16, pp. 267–278, 2018.

[41] K. Sarpatwar, R. Vaculin, H. Min, G. Su, T. Heath, G. Ganapavarapu, and D. Dillenberger, "Towards Enabling Trusted Artificial Intelligence via Blockchain," *Policy-Based Autonomic Data Governance Lecture Notes in Computer Science*, pp. 137–153, 2019.

[42] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, 2016, August. "Medrec: Using blockchain for medical data access and permission management," In *2016 2nd International Conference on Open and Big Data (OBD)* (pp. 25-30). IEEE.

Jennath H.S.

Jennath received a B.Tech in Electronics and Communication Engineering from University of Calicut in 2005. She worked as Technology Lead in IT companies like Infosys and IBS Software Services after her graduation. She did her M.Tech in Wireless Networks and Application from Amrita Vishwa Vidyapeetham in 2015. She is currently pursuing her Ph.D. from Cochin University of Science and Technology and her research centre is Indian Institute of Information Technology and Management – Kerala (IIITM-K). Her research interests include Vehicular Communication, Wireless Networks, Machine Learning and Blockchain.



Anoop V.S.

Anoop has completed his research leading to Ph.D. in Computer Science with specialization in Artificial Intelligence, from Indian Institute of Information Technology and Management – Kerala (IIITM-K) under Cochin University of Science and Technology. He has also completed a Master of Philosophy (M.Phil.) with specialization in Artificial Intelligence from IIITM-K and Master of Computer Applications from IGNOU, New Delhi. Anoop is currently associated with Kerala Blockchain Academy in the role of a Senior Scientist and Head of Research &Training. Prior joining KBA, Anoop has worked with Etisalat and Ethihad Airways as part of Dubai Future Accelerator and Fika Labs Program respectively in the United Arab Emirates. He is an experienced Software Engineer with over six years experience and also served IGNOU as a faculty member for its School of Computer and Information Sciences. Anoop's primary research interests include Applied Artificial Intelligence using Text Mining, Natural Language Processing (NLP) and Information Retrieval and Blockchain. He has several publications in his credit including international and national journals, book chapters and conference proceedings in major venues of Artificial Intelligence and Knowledge Management.

Asharaf S.

Dr. Asharaf S is a Professor at Indian Institute of Information Technology and Management – Kerala. He is also serving as a Professor-in-Charge for Kerala Blockchain Academy (KBA) and also a visiting faculty in Indian Institute of Space Science and Technology, Trivandrum and as a Mentor in Kerala Startup Mission. He received his Ph.D. and Master of Engineering degrees in Computer Science from Indian Institute of Science, Bangalore. He graduated in Computer Engineering from Cochin University of Science and Technology. After his PhD he has worked with America Online (AOL) and IIM Kozhikode. He is a recipient of IBM outstanding PhD student award 2006, IBM Shared University Research Grant, 2015 and IBM Open Science Collaboration Programme grant, 2017. He has published three books and more than 30 research papers in international journals and conferences. His areas of interest include technologies and business models related to data engineering, machine learning, information retrieval and blockchains.

Blockverse: A Cloud Blockchain-based Platform for Tracking in Affiliate Systems

A. Baldominos[1,2]*, J. L. López-Sánchez[2], M. Acevedo-Aguilar[2]

[1] Computer Science Department, Universidad Carlos III de Madrid (Spain)
[2] Blockverse Solutions S.L., Madrid (Spain)

Received 6 April 2020 | Accepted 14 April 2020 | Published 15 June 2020

Abstract

Affiliate systems are a crucial piece of today's online advertising. In affiliate systems, web traffic is directed from certain sites displaying ads to the websites of those company whose products or services are advertised. The way in which these ads are monetized is diverse and can respond to different models. In many cases, affiliates establish a cost based on impressions (displays of the ad) or on clicks. However, more intricate models are becoming widespread, such as the cost per action, where the affiliate incomes are due to the users performing certain actions in the target website. In particular, in the world of iGaming, it is frequent that affiliates charges are based on registrations, deposits or money lost on bets. In this scenario, Blockverse is a tool whose objective is to record transactions occurring in affiliate systems at large scale, using a permissioned blockchain implemented atop state-of-the-art cloud technology. Additionally, the system will be able to execute smart deals that generate income for affiliates based on the agreed conditions, and to provide real-time analytics in the context of the affiliate system.

Keywords

Blockchain, Cloud Computing, Databases, Affiliate Systems, Analytics.

DOI: 10.9781/ijimai.2020.06.001

I. Introduction

AFFILIATE systems play a key role in today's online advertising scheme, as their presence is key to gain web traffic to those websites whose products or services are being advertised.

Although the methods and agreements by which affiliate systems are governed are quite diverse and strongly depend on the industry whose products or services are advertised, the idea behind all of them is quite simple: affiliate systems *sell* the product of another company (the advertiser) at an agreed fee or commission.

Now, this *sale* can take place in different ways. In this paper, we will focus in a certain type of affiliate systems which direct web traffic to the advertiser, with the latter performing the final sale. In other words, the sale itself is performed as usual, except for the fact that the customer did not arrive organically, but forwarded by another company, which must receive the agreed fee.

Again, the deals involving how these fees are determined can vary significantly between industries and products. One example of such fees can be found in the *Amazon Associates* affiliate program [1], where affiliates can sell products of Amazon.com for a certain commission. As of March 2020, this commission is determined as a percentage based on the category of products (e.g. 10% in luxury and beauty, 4.5% in physical books, etc.) or as a fixed fee for the registration in some services (e.g. $3 for registering into Amazon Prime free trial).

In other cases, commission fees are subject to specific agreements resulting from negotiations between the partners (the affiliate and the advertiser), that could be re-negotiated in the future.

* Corresponding author.
E-mail address: abaldomi@inf.uc3m.es

In this paper, we will put the focus on the industry of iGaming. In this setting, we will call *operators* to the advertisers, since they are expected to have a *license to operate* in the iGaming market. Examples of operators can be online casinos or bet platforms.

In iGaming, the deals governing how affiliates perceive money can be quite intricate and complex. For example, the affiliate may perceive a fixed fee if a customer registers in the operator's platform. Additional fees, either fixed or variable, can be agreed if the customer deposits money or losses it in a bet.

Additionally, there might exist several agreements established between an affiliate and an operator, based on *how* the customer arrived into the operator's platform. For example, certain advertising sites can be paid at a better rate or, in some cases, the fees can be determined based on the size of the ad, etc.

This complex scheme makes it critical to carry out secure tracking of customers, from the moment they arrive at the advertising to their steps in the operator's platform. The identification of customers must be unequivocal in order for affiliates to perceive agreed fees correctly.

In this scenario, it is frequent that a certain degree of distrust arises. In many cases, a chain of intermediaries exists between the affiliate and the operator in order to alleviate some of this distrust, therefore adding extra costs to the entire advertising chain.

In a setting like this, a Blockchain-based system can serve as a technical solution for securely and trustily logging the interactions and communications between affiliates and operators, without the need for additional intermediaries, therefore reducing costs while keeping trust.

The Blockchain would be used as an immutable ledger storing all of the transactions taking place in the advertising scheme, therefore enabling the partners to performed advanced analytics, while at the same time would allow the establishment of custom smart deals

between affiliates and operators in order to govern how exchanges of traffic and money are carried out.

In this paper we are devising Blockverse, an online platform based on Blockchain for tracking the transactions between affiliates and operators in the iGaming industry. In particular, the paper explains technical decisions made for allowing Blockverse to scale and to provide a solution to the problem of tracking customers, traffic and transactions.

The paper is structured as follows: section II presents some basic concepts and notation that will be useful to understand the remainder of the proposal, aimed at people not familiarized with the industry or with affiliate platforms. Later, section III presents the current state of the art, summarizing related works where the Blockchain is being studied as a solution for similar problems and domains.

The Blockverse platform is described in section IV, illustrating its user experience and discussing how communication between affiliates and operators is established. This is followed in section V by a technical discussion of how the system is designed to scale to large amounts of traffic. Finally, section VI presents some conclusive remarks of this paper and poses interesting lines for future research.

II. Basic Concepts and Terminology

In this section, we introduce some basic notation that will be useful to understand the remainder of the proposal.

A. Concepts Related to Affiliates

An affiliate is an entity that forwards traffic to the advertiser, which in the iGaming industry we also refer to as the *operator*. The affiliate is interested in that this traffic is of quality, in the sense that it leads to a higher rate of conversions (specific actions in the operator's platform), since this will often result in higher revenues.

The following concepts are closely related to the role of the affiliate in the advertising scheme:

- Money site: it is a webpage designed to redirect traffic to the operator's platform. This webpage might contain ads in the form of banners, videos or other form of sponsored content with links to the operator's site.
- Asset: it is a resource which is placed in money sites and serves for the purpose of forwarding the customer to the operator's platform, such as a clickable banner.
- Click event: it is an event initiated by the customer (an online users) that results into a redirection of traffic to the operator's platform, such as clicking an asset.

B. Concepts Related to Operators

An operator is a company which holds a license to operate in the iGaming industry, such as a casino or betting platform. Due to the nature of this industry, there are strong regulatory platforms to ensure that companies provide a reliable service to customers, and companies must adhere to these regulations to operate.

The following concepts are closely related to the role of operators in the advertising scheme:

- Platform landing page: it is the website in which a customer *lands* after performing a click event. This website will often present the operator's features and allows the customer to register, in some cases providing customers with some offers or other advantages for promoting their enrolment.
- Brand: in some cases, operators might present to their customers different *brands* or commercial names. For example, a single operator might have one brand for an online casino and a different one for sports bets.
- Register event: it is an event initiated by customers that results in their enrolment in the platform.
- Deposit event: it is an event initiated by customers that leads to the deposit in their platform's account of a certain amount of money. This money can later be used for betting and can be incremented or decremented based on the outcome of bets. This amount can also be decreased when customers transfer money to their bank account.
- Bet event: it is an event initiated by customers that consist in betting a certain amount of previously deposited money. This bet can take place in an online roulette, a poker game, a soccer match, etc.
- Win/lose: it is the outcome of a bet event. This outcome might be known immediately after the bet (for example, in an online roulette) or after some time (for example, when a soccer match has concluded). A *win* outcome means that the customer wins some money beyond the betted amount, therefore increasing the money in the deposit. A *lose* outcome means that the customer loses money, which might be all of the betted amount. It is worth noticing that the operator earns money when the customer loses a bet.

C. Users vs. Customers

So far, we have been using the term *customer* to refer to an online user during the whole interaction process, including when he/she visits a money site. This terminology might be imprecise and lead to confusion, since a user might not be considered as such before registration in the operator's platform, or even before deposit.

Properly speaking, a user is not considered a customer until a certain event occurs, which is known as the *conversion*. Most often, a user that lands into the operator's platform (e.g. after a click event) is not yet considered a customer, although has potential to become into one.

In some cases, the conversion event is the register event itself. In other cases, a user might not be considered to convert into a customer until a deposit takes place. This can happen, for instance, in some platforms with a small ratio of deposits/registrations.

Anyhow, the difference between users and customers is not a critical issue in this paper. Later, we will see how the Blockverse platform has chosen a term to denote both users and customers regardless of when conversion takes place.

D. Concepts Related to the Market

Besides those concepts that are exclusive to affiliates or operators, there are some important terms that are relevant in the interaction that arises between them:

- Partnership: it is the agreement between an affiliate and an operator to establish certain deals for the former to redirect traffic to the latter.
- Transaction: it is any exchange of information between the two partners. Additionally, a transaction can also take place between the customer and the operator, but even in this case the affiliate will be considered a third-party in the transaction.
- Deal: it is an agreement of conditions for the affiliate to perceive some revenue due to customers generated by their traffic. These conditions can be very complex and can depend on aspects such as the money site, the asset generating the traffic, etc.

In section IV we will provide more specific details on the different types of deals that are supported in the Blockverse platform.

III. State of the Art

The blockchain is a concept generally acknowledged to arise in the original white paper describing the Bitcoin electronic currency by

Nakamoto [2]. It was originally developed as an auxiliary technology to support such cryptocurrency, providing the security mechanisms guaranteed by Bitcoin and enabling the mining procedure.

Later on, larger possibilities of the blockchain technology as Buterin described its ability for storing and running decentralized apps and smart contracts [3], which eventually led to the development of the Ethereum platform and the Ether cryptocurrency.

Since the mid-2010s, the blockchain has emerged as an interesting technology enabling a large variety of applications. A survey by Zheng *et al.* [4] in 2018 reveals a wide spectrum of applications to fields as diverse as finance, security and privacy, internet-of-things, reputation systems and public and social service. More recently, Baldominos *et al.* [5] have suggested mechanisms for evolving *proof-of-work* into mechanisms allowing for the distributed training of artificial intelligence applications.

An interesting feature of the blockchain is that it is often described as a mechanism for shedding trust into ecosystems where different parties interact with each other. For example, Blasingame [6] contributed an article to Forbes describing blockchain technologies as "the future of trust". Anjum *et al.* [7] described how blockchain could bring compliance and trust to different applications outside the scope of finance and banking. Also, Hawlitschek *et al.* [8] described how the blockchain is able to replace trust to a certain extent in the context of sharing economy applications.

In this paper, we put the focus into an application of blockchain to online advertising and affiliate systems. The literature coverage of this topic is rather narrow. A broad analysis of the impact of blockchain technologies on marketing was published by Ertemel in 2018 [9].

A more relevant work was presented by Pärssinen *et al.* [10], where they describe the potential of blockchain to online advertising. In this paper, authors enumerate some of the typical problems of the industry: large numbers of intermediaries adding costs and exploiting users' data, and fraud in the advertising chain. They end up concluding that further research is required to understand whether successful applications of blockchain to online advertising can be implemented. The issue of fraud has been further studied by Kshetri and Voas [11], suggesting that blockchain has a potential for tackling fraud, despite of some challenges regarding its application.

To the best of our knowledge, there are not published works describing applications of blockchain to affiliate systems. However, an early attempt was suggested by Weijer in his Master's thesis [12]. In his work, Weijer identifies some opportunities for blockchain in the scope of affiliate marketing and suggests a design for such a system, concluding that blockchain could improve transparency in the system.

Finally, although previous implementations of a system like this in the industry are not known by the authors to exist, a patent registered by IBM in 2016 [13] settles the foundations for a transactions tracking system bases on blockchain.

IV. The Blockverse Platform

In this paper, we are proposing and describing the implementation of a Blockchain-based platform for tracking the transactions taking place between affiliates and operators in an iGaming setting.

The rationale behind this platform is to enable a secure and immutable exchange of information between the partners, without the need of third-party intermediaries to preserve the trust of the system.

Behind the blockchain database used to store transactions, the platform is designed to enable an end-to-end interaction between the partners. This includes adding new partners to their network, establishing new deals, exploring transactions in real time, and checking out summary analytics that are automatically computed by the system. Therefore, Blockverse aims to be an integrated solution combining tracking of operations and enforcement of the execution of deals, as well as providing a business analytics tool.

This section provides a logic description of the key elements comprised in the Blockverse platform, whereas section V will delve into the technical details of the solution, putting the focus on the issue of scalability. Finally, section VI will provide some conclusive remarks about the Blockverse platform, and outline some lines of future work and improvements.

A. The Blockchain

The Blockchain is the database infrastructure used to store all the transactions taking place in Blockverse. In the most fundamental terms, a blockchain is no more than a *chain of blocks*, with each block storing certain information.

The blockchain technology; however, is famous for constituting the foundation of several well-known cryptocurrencies, such as Bitcoin or Ethereum. Decentralized blockchain networks such as those supporting these currencies provide mechanisms for ensuring integrity and security of the data via cryptographic mechanisms that take place during the process known as *mining*.

However, some of these cryptographic mechanisms can be used in a different type of blockchain network to still preserve data integrity. In particular, Blockverse relies on a centralized permissioned private network which implements security mechanisms to avoid manipulation of the ledger. These features of the blockchain technology underlying Blockverse have the following meaning:

- Centralized: the database is stored in a centralized system, as opposed to a distributed system composed of a network of nodes carrying out the mining process.
- Private: the database is not public, meaning that not anyone has access to its contents.
- Permissioned: the database will provide selective access of the contents of the blockchain to users based on their permissions, meaning that a given user will often have a very restricted view of the database (such as those transactions in which he/she is involved).

The blockchain technology used for Blockverse is the result of a custom implementation, which is strongly influenced by the Bitcoin technology. In this implementation, transactions are stored in the database and are periodically included into a block. Since the blockchain is centralized, a distributed mining system does not make sense. However, Blockverse has incorporated an automated mining system which confirms blocks after ensuring that all transactions located within them are valid and consistent with the rules governing the market.

As in Bitcoin, a cryptographic mechanism is implemented to ensure the integrity of the chain. In particular, each block is hashed using SHA3-512 and the hash of a block is used as an element of the following block (and included in the latter's hash). Therefore, trying to break the integrity of the blockchain will raise an alarm, as hashes would not match their expected values.

B. Trackers

In section II we explained the difference between a user and a customer. This distinction can be hard to do because it will often depend on the definition given by operators, when in some cases a customer will be a registered user and, in others, the user will need to deposit a certain amount to become a customer.

In any case, Blockverse has introduced the concept of *tracker* to allow the system to remain agnostic about the real nature of users and

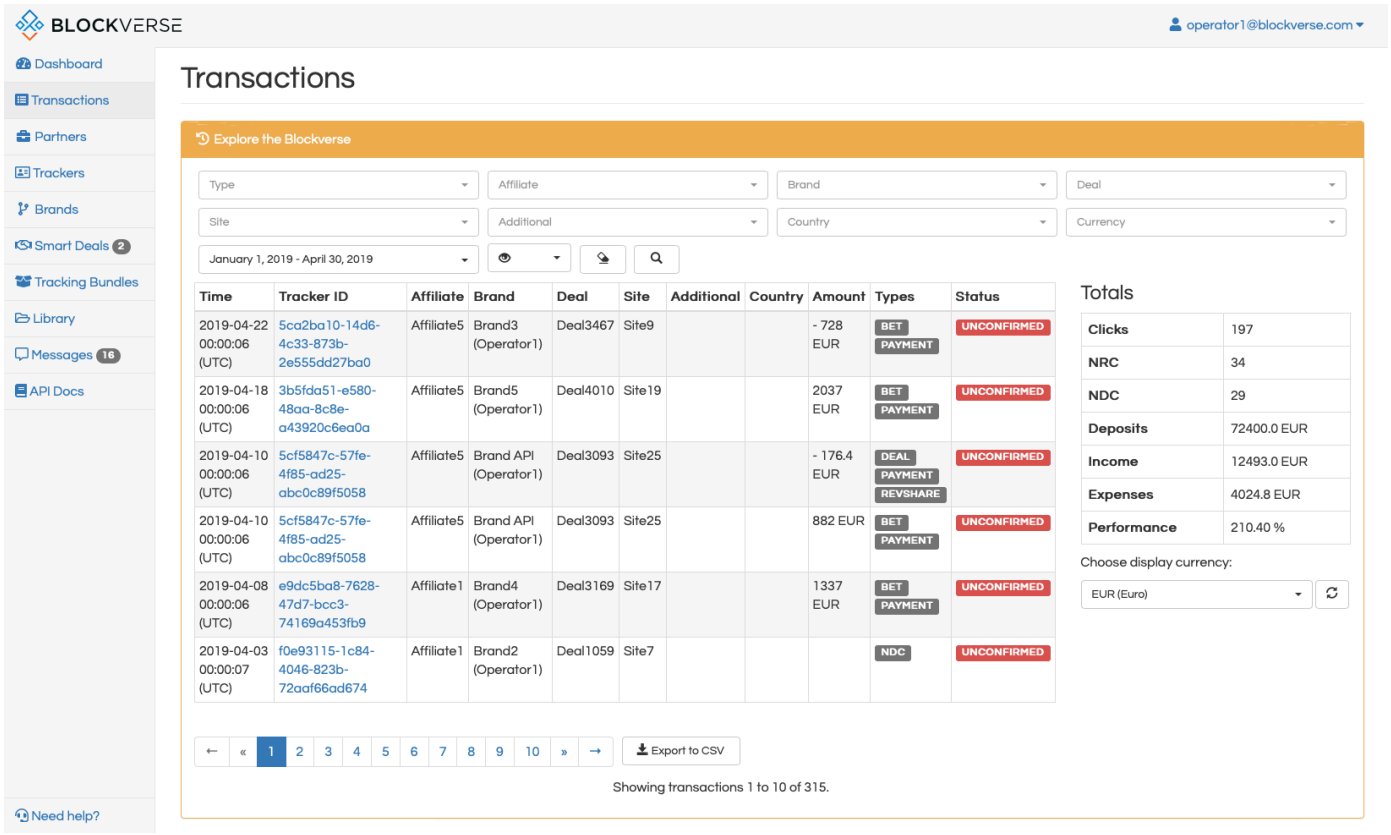


Fig. 1. Transactions screen in Blockverse. In this view, both affiliates and operators can see the list of transaction, applying filters as required. Custom fields are shown as columns in the table and are displayed based on the permissions of the logged profile. A section of *totals* is shown in the right side, providing some metrics of interest for the filtered data. All transactions can be exported to a CSV file, so affiliates and operators can perform more complex queries or data analysis processes as required.

customers. A tracker is an entity that needs to be tracked, regardless of whether the conversion takes place.

A tracker is generated in Blockverse as soon as a click event takes place. In that moment, the user is forwarded to the operator's platform and the unique identifier of the tracker freshly generated is provided to the operator as part of the redirection process. During the tracker creation, additional information might be stored about how this tracker emerged, for example, the money site and asset that led to the generation of the tracker, etc. This will be useful to choose which smart deals (specific business logic affecting the revenue of affiliates) will govern the actions performed by such tracker, if any.

After landing in the operator's platform, the operator will be in charge of linking the tracker identifier to any of the operations performed by the tracker, and particularly, reporting to Blockverse the register, deposit, bet and lose/win events. Such events will be used with two purposes: (1) triggering smart deals that generate revenue to affiliates based on the tracker actions and (2) providing accurate business analytics and key performance indicators to both affiliates and operators about the behavior of trackers.

Even if the user does not perform any additional action beyond landing in the operators' platform, Blockverse will create the tracker, since it can be used by affiliates to get a summary of the performance of their different money sites and assets.

Finally, trackers in Blockverse have an associated key-value data store, which can be used by partners to provide additional information about trackers. Interestingly, these data are stored outside of the blockchain; instead, the blockchain only stores a unique identifier of trackers in the transactions in which they are involved. This means that this data store can be used to fill trackers with information known about them (name, email, etc.) allowing GDPR compliance since the data could be modified or removed if requested by the user.

C. Transactions

In Blockverse, a transaction is any kind of interaction between partners that is stored in the blockchain. Most transactions will always involve three parties: an affiliate, an operator, and a tracker.

Transactions in Blockverse are labelled with a type and can have associated data in the form of a key-value store.

While the blockchain implementation available in Blockverse provides support for transactions of any type, the transaction types are currently restricted to the following: *new* (for click events), nrc (standing for *new registered customer,* it used for register events), *deposit*, *ndc* (standing for *new deposit customer*, it is recorded the first time a tracker completes a deposit event), *bet* and *payment*. Affiliate platforms can record transactions of type *new*, whereas operators can record any of the other.

Payment transactions can have two different sources. On the one hand, they can be reported by the operator when a win/lose event takes place, indicating which are the source and the target of the payment.

On the other, *pay* transactions can be generated automatically by smart deals as a result of commission fees due to the events performed by the tracker. These *pay* events represent income of the affiliate that should be invoiced to the operator at the end of the period.

Regarding the data store associated with a transaction, it is able to store any kind of data. These data can be introduced either by the affiliate or by the operator when a transaction is created.

Data items associated with a transaction are permissioned as well, meaning that the owner of the transaction (the partner that records it)

can decide whether a piece of data is private or public. By *public*, we mean that the other partner has access to it. As a result, private data can be useful for an affiliate or an operator to store information internal to their business logic or that is confidential.

An example of how transactions are displayed in the Blockverse website is displayed in Fig. 1. It can be seen how different filters can be applied and some summarizing metrics are computed for the specified period and conditions.

D. Smart Deals

Smart deals in Blockverse constitute a mechanism that mimics the idea of smart contracts that is present in some blockchains, such as Ethereum. Unlike smart contracts, Blockverse deals are not stored inthe blockchain, but are stored as rules in the application, that are checked for execution every time new transactions are recorded.

Smart deals are the technology enabling the computation of amounts of money that operators shall pay affiliates for their services. The contents of smart deals are negotiated by partners and are they are recorded as application logic once terms have been agreed.

As of 2020, four types of smart deals can be established:

- Flat fee: it allows to set up a monthly fee that is paid from operators to affiliates regardless of traffic. This is a special kind of smart deal, since it is not triggered by transactions recorded in the blockchain, but on a timely basis.
- CPA: it stands for *cost per action* and allows to pay a certain fixed amount when an action occur. At this moment, the only supported action is NDC; i.e., when a customer makes a first deposit of money in the operator's platform. In this case, partners should agree on the CPA amount (the amount that will be paid to the affiliate if the action takes place) and can also establish a minimum deposit threshold, meaning that the payment will only take place if the users deposit at least that amount of money in the operator's platform.
- Revenue share: it allows to pay a variable amount of money based on the operator's revenue coming from lost bets. This amount is agreed as a percentage.
- Hybrid: it is a combination of CPA and revenue share, allowing to define both a fixed amount based on NDC actions and a variable amount based on revenue.

When trackers are created, they are already linked to a certain smart deal of CPA, revenue share or hybrid type. The way in which deals are linked to trackers will be explained later in this section. This means that affiliates and operators can apply different deals based on the source of traffic; for example, certain privileged positions (e.g. larger ads) can be subject to more interesting deals for the affiliate.

For all of these deals, partners can agree on a time threshold (specified as days, weeks or months), during which the deal will remain active for each tracker. This allows operators to decide for how long they will be paying affiliates for the revenue of a tracker or, in the case of CPA deals, after how much time they will not pay the affiliate for a new deposit. In either case, the period of time starts counting on the creation of the tracker.

Additionally, deals can be set up with an expiration date. This value is tracker-agnostic, meaning that after that date the deal will no longer be triggered, regardless of how many trackers are associated to it.

Finally, Blockverse provides an interface to create and negotiate deals. Smart deals can be created by either affiliates or operators, and they have to choose the partner to whom the deal will be established. After creation, the deal's status is "awaiting approval", and the recipient partner will be notified that a new deal is pending for verification. The screen available in Blockverse for creating a new smart deal is shown in Fig. 2.

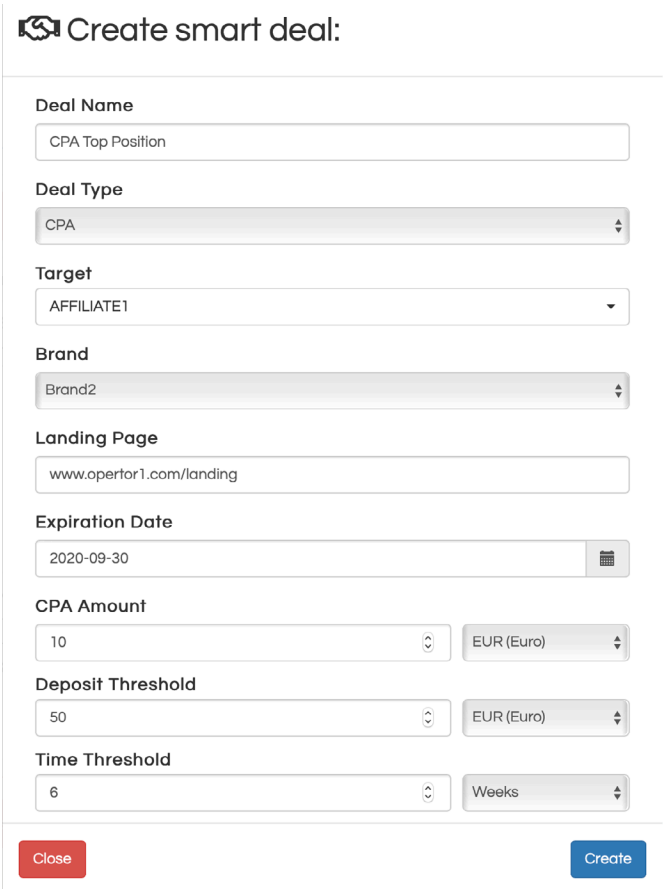


Fig. 2. Screen for setting up a new smart deal in Blockverse. In this example, the view for a new CPA deal is shown, although logic is similar for flat fee and revenue share deals. Once the deal is created, the partner must approve it.

At this moment, the recipient partner can decide to accept the deal, deny it or make further modifications to it. If the deal is approved, it will be set up as "active" and will start working. If the deal is "denied", it will be erased from the platform, and the issuer will be notified. If the deal is modified, it is returned to the issuer for its approval.

Active deals can be modified or cancelled, but in any of the cases, a request is issued to the other partner, and changes are not applied in the deal's logic unless the other party accepts the changes. This process ensures that deals are never modified unilaterally.

E. API and Tracking Bundles

Blockverse allows both affiliates and operators to record transactions using a RESTful HTTP API. All of the business logic, including the workflow of smart deals and retrieval of analytics, can be handled through this API.

A typical use case would be the following: an affiliate and an operator agree on a smart deal (either using the Blockverse interface or the API). Once the deal is created, it can be referred to using its unique identifier. The affiliate will set up an ad in a money site of its property, that will redirect traffic to the operator.

When the ad is clicked, the affiliate will call the Blockverse API to create a new tracker linked to the previously created deal. The Blockverse API will return a unique identifier for the tracker, and the affiliate will provide the operator with this identifier. In subsequent actions of the tracker, the operator will call the Blockverse API to record such transactions, communicating the tracker identifier. During each transaction, Blockverse will automatically check if the smart deal associated with the tracker involved in the transaction should be triggered, doing so if required.

To ease this whole process, Blockverse also provides a mechanism known as "tracking bundles", which is a method to simplify tracker creation on the side of the affiliate.

When creating a tracking bundle, the affiliate will need to specify the operator and operator's brand, the smart deal involved and the landing page on the operator's platform. Moreover, the affiliate can set up custom fields in the form of pairs of keys and values, that will be stored along with the transactions resulting from the interaction with that tracking bundle.

Operators can also see tracking bundles referred to them and add additional custom fields. In either case, these fields visibility can be adjusted as public or private, and in the latter case, only the owner will be able to see them. The interface for creating new tracking bundles in Blockverse can be found in Fig. 3.

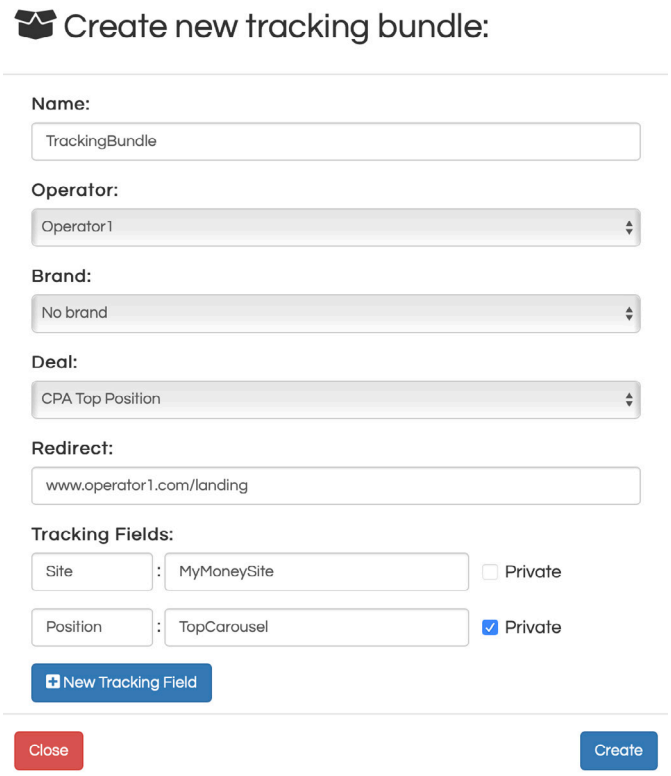


Fig. 3. Screen for setting up a new tracking bundle. Both affiliates and operators can set up custom fields and adjust their visibility.

After creating a tracking bundle, Blockverse will provide an URL that can be called directly by the affiliate when the *click* event takes place (for example, as the target URL of an ad). When this URL is called, Blockverse will automatically create the tracker and redirect the user to the operator's platform (in particular, to the URL specified in the "Redirect URL" field in the bundle creation form, see Fig. 3). The identifier of the freshly created tracker will be passed to this URL as a query parameter.

Additionally, the Blockverse platform will also display cURL and Python code required to record the *click* transaction with the data specified in the tracking bundle, to ease API integration.

F. Revokation of Transactions

Since all the transactions recorded in Blockverse are stored in a blockchain, past transactions cannot be deleted, even if they are known to be wrong (e.g., an operator records a transaction incorrectly).

To tackle this problem, Blockverse enables a mechanism to flag transactions as "revoked" that do not break the blockchain integrity. This is carried out using a separate list of pointers to those transactions that are revoked.

Additionally, operators are given the possibility to revoke trackers. This mechanism is implemented to prevent potential fraud; e.g., affiliates redirecting traffic using a different deal than the one agreed to obtain higher revenues. By revoking a tracker, all past and future transactions associated to it are revoked, including those that constitute a source of revenue for the affiliate. New transactions will be still recorded, although they will be automatically revoked, and smart deals will no longer be executed.

G. Dashboard and Business Analytics

So far, we have described the features of Blockverse regarding the logic by which transactions are recorded and smart deals are executed.

Beyond this functionality, Blockverse provides a dashboard that displays relevant information to both affiliates and operators, that can be used to better understand their sources of traffic and revenue and take direct actions accordingly.

The top part of the dashboard provides the following information:

- Total clicks in the current month.
- Total NRCs (new registered customers) in the current month.
- Total NDCs (new deposit customers) in the current month.
- Total deposits in the current month.
- Total revenue in the current month.

All of these data are provided along with the information for those metrics in the same period of the previous month, allowing partners to better understand their current performance.

Additionally, some graphics are provided to display in a single image some relevant information. In the case of affiliates, the following graphics are displayed:

- A cumulative bar chart showing total revenue by position. Only the top ten positions are shown, and revenue is split based on whether it comes from CPA or revenue share.
- A monthly timeline plotting the evolution of the clicks to NRC ratio for the top five operators. This allows affiliates to know which operators are converting more traffic into registered users.
- A monthly timeline plotting the evolution of the clicks to NDC ratio for the top five operators. This allows affiliates to know which operators are providing better conversion rates and, since NDC is directly related to income from CPA smart deals, it enables affiliates to be at a better position to negotiate new deals.

In the case of operators, the first chart (revenue by position), is replaced by a bar chart plotting the traffic from affiliates, measured as *click* events for the top five affiliates and including the monthly evolution. The other two charts (clicks to NRC and clicks to NDC ratios respectively) are equally shown but referring to the top five affiliates partnered with the operator. This enables operators to identify which affiliates are sending traffic of higher quality; i.e., more easily translated into conversions.

Finally, as it could be seen in Fig. 1, a summary of information can be shown for any set of transactions, applying as many filters as the partner wants. Filters can also be set on custom attributes included in a transaction; therefore, enabling complex analytics. Blockverse allows partners to download this information as a CSV file that can be loaded into external business analytics tools.

V. System Scalability

Due to its nature, Blockverse is a system that must be designed with scalability in mind, since it needs to adapt to large amounts of

incoming traffic (from all affiliates and operators subscribed in the platform). It is important that the system is responding at all times, properly recording all the transactions. Otherwise, lost transactions would not only mean a loss of trust in the analytics provided by the system but could also lead to potential losses of revenue for affiliates.

To prevent this from happening, the system is designed with a robust and reliable cloud-based architecture that facilitates scalability of the system as traffic increases.

The whole Blockverse system is deployed in Amazon Web Services. Fig. 4 summarizes the current cloud architecture.

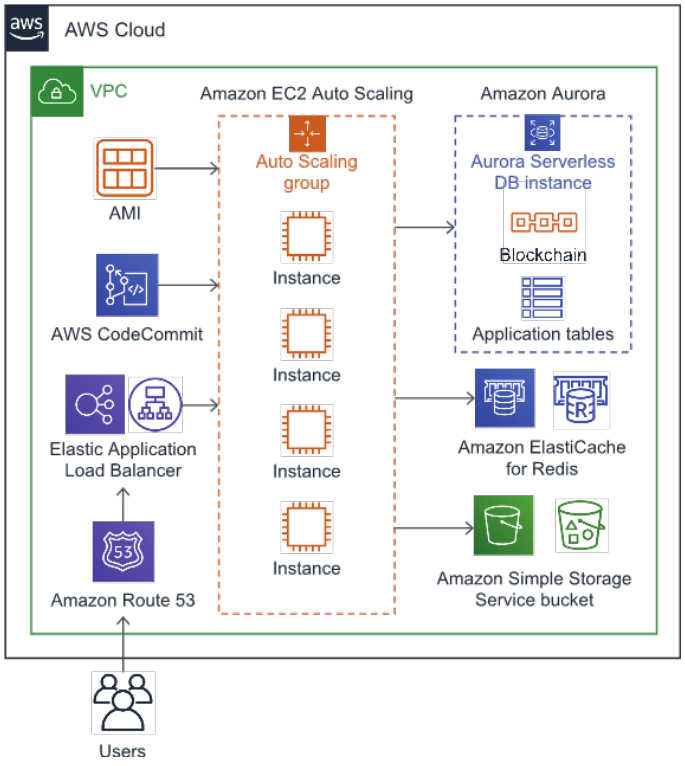


Fig. 4. Diagram of the cloud architecture of the Blockverse platform, representing how the different cloud services interact with each other.

The production environment is deployed within an autoscaling group, a mechanism that allows to provision new instances or destroy them based on traffic or CPU usage. Every time a new instance is provisioned, it launches the latest version of the Blockverse AMI (Amazon image) and checks out the most recent software release from CodeCommit (Amazon's source code repository) before starting to attend requests. Additionally, new software releases are rolled out to instances in the autoscaling group progressively, in order to prevent service interruption.

Data is stored in a relational database which is deployed in an Aurora serverless instance within Amazon RDS (relational database service). This database contains both the application tables and the blockchain itself. Aurora serverless enables automatic scalability based on traffic, provisioning additional compute units as traffic increases. Replication and automated backups are automatically managed by RDS.

Additionally, an in-memory Redis-based database is used for managing session information in Blockverse, using Amazon ElastiCache service. Also, S3 (simple storage service) buckets allow to store certain objects that are exchanged between partners, such as assets that they can upload to a shared library.

An application load balancer is in charge of redirecting traffic to the different instances in the autoscaling group. These two elements are configured together so that the load balancer is aware of which instances are available at each time. Finally, Route 53 is used to configure DNS entries for resolving domain names to the load balancer, and SSL certificates are also managed by Amazon to ensure secure communication between Blockverse and the users.

All the system is deployed within a VPC (virtual private cloud) and security groups have been configured to allow only authorized communications between the different services.

VI. Conclusions and Future Work

In this paper we have presented Blockverse, a blockchain-based platform for tracking transactions in affiliate systems. At its current implementation, the system is deployed in the iGaming industry. In this industry, we can distinguish two different parties: operators (companies licensed to operate in the market) and affiliates, who redirect paid traffic to operators.

This industry is complex inasmuch as operators and affiliates can establish very complex agreements on how the former will pay the latter based on the received traffic. Generally, operators will pay affiliates based on actions (e.g., the user deposits some amount of money in the platform) or as a revenue share of the income generated by the user. However, the specifics of these agreements can be quite intricate, depending on the type of ad, the location of the banner, etc.

The Blockverse platform provides means for affiliates to record the traffic they generate to operators, and for operators to record the actions performed by users in their platform. Blockverse keeps track (using a mechanism conveniently called *trackers*) of how affiliate traffic converts into registrations, deposits or bets. Both partners can record custom fields associated to each transaction, that can later be used to perform advanced filtering and analytics.

Additionally, partners can use Blockverse to negotiate and establish *smart deals*, which rules determining the conditions in which affiliates will perceive money for traffic redirected to the operator's platform. Negotiation of deals has an established workflow that ensures that both partners agree on the final details. Trackers can be assigned to different deals, based on the different ways in which affiliates get traffic to operators, such as different money sites or positions.

The kind of industry in which Blockverse takes place imposes remarkable requirements when it comes to availability and scalability. For example, missing some API requests can have a direct impact in the revenue perceived by affiliates. To tackle this, we have designed a system which is entirely cloud-based, relying on different AWS tools to ensure that scalability is guaranteed, including serverless deployments of the database auto-scaling setups assisted by load balancers for the computing infrastructure.

In the shortcoming future, Blockverse functionality and its underlying architecture can be further improved. Regarding functionality, a broader consideration of CPA smart deals could be interesting besides NDC events. For example, *register* events could constitute another action which is a source of revenue for the affiliate, and which at this moment is not supported.

Regarding the Blockverse deployment on cloud, it could be interested to migrate the blockchain from its current SQL implementation to QLDB (Quantum Ledger Database), a recent serviced introduced by Amazon [14] that provides a quite natural mechanism to provide a transactions ledger mimicking the structure of a blockchain and without need to implement our own verification and integrity mechanisms.

These future improvements are already considered in the Blockverse roadmap as future development.

References

[1] Amazon, "Amazon.com Associates Central," [Online]. Available: https://affiliate-program.amazon.com/help/node/topic/GRXPHT8U84RAYDXZ. [Accessed 20 03 2020].

[2] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2009. Available: https://bitcoin.org/bitcoin.pdf [Accessed 12 06 2020].

[3] V. Buterin, "A next-generation smart contract and decentralized application platform," Ethereum White Paper. 2014. Available: https://ethereum.org/whitepaper/ [Accessed 12 06 2020].

[4] Z. Zheng, S. Xie, H.-N. Dai, X. Chen and H. Wang, "Blockchain challenges and opportunities: A survey," *International Journal of Web and Grid Services*, vol. 14, no. 4, p. 352–375, 2018.

[5] A. Baldominos and Y. Saez, "Coin.AI: A Proof-of-Useful-Work Scheme for Blockchain-Based Distributed Deep Learning," *Entropy*, vol. 21, no. 8, 2019.

[6] J. Blasingame, "Blockchain Isn't The End Of Trust, It's The Future Of Trust," *Forbes*, 07 Jul 2019.

[7] A. Ashiq, M. Sporny and A. Sill, "Blockchain Standards for Compliance and Trust," *IEEE Cloud Computing* , vol. 4, no. 4, p. 84–90, 2017.

[8] F. Hawlitschek, B. Notheisen and T. Teubner, "The limits of trust-free systems: A literature review on blockchain technology and trust in the sharing economy," *Electronic Commerce Research and Applications*, vol. 29, p. 50–63, 2018.

[9] A. V. Ertemel, "Implications of Blockchain Technology on Marketing," *Journal of International Trade, Logistics and Law*, vol. 4, no. 2, p. 35–44, 2018.

[10] M. Kotila, R. C. Rumin, A. Phansalkar, J. Manner and M. Pärssinen, "Is Blockchain Ready to Revolutionize Online Advertising?," IEEE Access, vol. 6, p. 54884–54899, 2018.

[11] N. Kshetri and J. Voas, "Online Advertising Fraud," *Computer*, vol. 52, no. 1, p. 58–61, 2019.

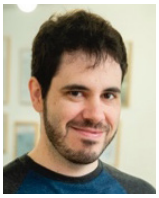
[12] M. Weijer, "Providing Trust in Affiliate Marketing through Blockchain Technology," Utrecht University, 2019.

[13] M. J. Bordash, M. J. Hudson and C.-H. Wong, "Tracking transactions through a blockchain". USA Patent US20180053161A1, 2016.

[14] Amazon, "Amazon Quantum Ledger Database (QLDB)," [Online]. Available: https://aws.amazon.com/qldb/. [Accessed 06 04 2020].

A. Baldominos

A. Baldominos received his Ph.D. in computer science and technology by Universidad Carlos III de Madrid, Spain, in 2018, with his dissertation focusing on the automatic evolutionary design of deep and convolutional neural networks. He works as a Research Assistant in Universidad Carlos III de Madrid (Spain) and teaches different courses in private colleges, such as Universidad Internacional de la Rioja and CUNEF Universidad. He is also founder and CTO at Blockverse Solutions. Dr. Baldominos is member of AEPIA, the Spanish Association for Artificial Intelligence, and is a former member of IEEE, ACM and AAAI.



J. L. López-Sánchez

J.L. López-Sánchez received his B.Sc. in computer science by Universidad Carlos III de Madrid, Spain, in 2014, with a minor in artificial intelligence. He currently works as a software developer for the group Multimedia and Teaching Innovation at Universidad Carlos III de Madrid. He is also founder and CIO at Blockverse Solutions.

M. Acevedo-Aguilar

M. Acevedo-Aguilar received his B.Sc. in electronic engineering by Universidad Carlos III de Madrid, Spain, in 2012, focused on embedded software. He works as Project Manager and Product Owner at different startups. He is also an experienced BI Consultant as well as CEO and founder at Blockverse Solutions.

Efficient Method Based on Blockchain Ensuring Data Integrity Auditing with Deduplication in Cloud

Mohamed El Ghazouani[1], My Ahmed El Kiram[1], Latifa Er-Rajy[1], Yassine El Khanboubi[2]

[1] Cadi Ayyad University, Marrakesh (Morocco)
[2] Hassan II University, Casablanca (Morocco)

Received 30 April 2020 | Accepted 20 July 2020 | Published 6 August 2020

Abstract

With the rapid development of cloud storage, more and more cloud clients can store and access their data anytime, from anywhere and using any device. Data deduplication may be considered an excellent choice to ensure data storage efficiency. Although cloud technology offers many advantages for storage service, it also introduces security challenges, especially with regards to data integrity, which is one of the most critical elements in any system. A data owner should thus enable data integrity auditing mechanisms. Much research has recently been undertaken to deal with these issues. In this paper, we propose a novel blockchain-based method, which can preserve cloud data integrity checking with data deduplication. In our method, a mediator performs data deduplication on the client side, which permits a reduction in the amount of outsourced data and a decrease in the computation time and the bandwidth used between the enterprise and the cloud service provider. This method supports private and public auditability. Our method also ensures the confidentiality of a client's data against auditors during the auditing process.

Keywords

Auditing, Blockchain, Cloud Computing, Deduplication, Integrity.

DOI: 10.9781/ijimai.2020.08.001

I. Introduction

CLOUD computing is defined by the National Institute of Standards and Technology (NIST) as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable Computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [1]. Cloud computing provides a number of opportunities, such as enabling services to be used without any understanding of their infrastructure, and that data and services are stored remotely but are accessible from anywhere. This way of remote storage is the most important cloud service because it allows cloud users to store their data from local storage systems to the cloud. According to the NIST's classification, the four major patterns of cloud deployment are the private cloud, community cloud, public cloud, and hybrid cloud [1]. Cloud service models are classified as software as a service (SaaS), platform as a service (PaaS) or infrastructure as a service (IaaS).

Once the data is stored on the cloud service platform, the data owners lose control over it. Although this technology offers many advantages, it also introduces new security challenges, especially those related to the integrity of the data. Data integrity is one of the most critical elements in any system. To ensure the integrity of outsourced data, a data owner should enable auditing mechanisms. Auditing is a process of analysis and verification, performed by an internal or external auditor, with the aim of identifying the security vulnerabilities of a system. In our paper, we use the auditing process to check the data integrity of the outsourced data.

* Corresponding author.

E-mail addresses: mohamed.elghazouani63@gmail.com (M. El Ghazouani), kiram@uca.ac.ma (M. A. El Kiram), errajy.latifa@gmail.com (L. Er-Rajy), elkhanboubi.yassine@gmail.com (Y. El Khanboubi).

The second important requirement when storing user data is storage efficiency. Data deduplication is the best choice for ensuring data storage efficiency. Data deduplication (also called intelligent compression or single-instance storage) eliminates redundant data and keeps just one copy of each file before the transfer of the data to be saved in the cloud server (deduplication on the client side is called source-based deduplication), or after it is transferred (deduplication on the server side is also called target-based deduplication). This technique means that multiple copies of the same data are not stored, which allows a reduction in data volumes and thereby reduces storage overhead.

Several cloud data integrity auditing protocols have been proposed in the last few years. In these protocols, the basic system model describes the various entities and their participation in the system, and the threat model highlights threats to an owner's data.

A. System Model: Private and Public Auditing

Several schemes are based on a private auditing system, which means that the data owner who audits the integrity of his data. In this type of auditing system, there are just two entities; the data owner and the cloud storage service (CSS):

- Data owner: the proprietor of the data; he is dependent on the cloud service provider for the proper maintenance of the data.
- Cloud Storage Server (CSS): the cloud service provider, who provides space to store an owner's data.

Fig. 1 shows a private auditing system. This system model provides the data owner with authority only to interact with the CSS to audit

data integrity and conduct data structure operations on outsourced data, whilst the readers only have the authority to read data.

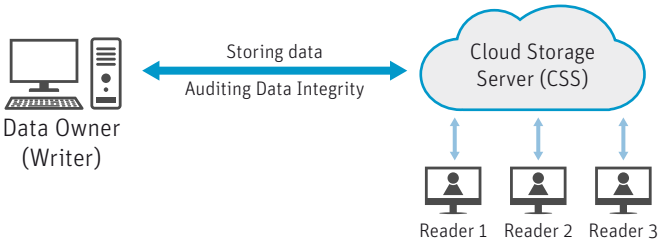


Fig. 1. Basic private auditing system.

Public auditing allows a third party to audit data integrity rather than the data owner. There are three entities in this type of system: the data owner, cloud storage server and a third party auditor (TPA). The TPA has the ability to access the services afforded by the CSS, and therefore, the data owner requests them to check the integrity of their data.

Fig. 2 shows a public auditing system. This system model provides authority only to a TPA to interact with the CSS to audit data integrity. The TPA can significantly alleviate the auditing costs of users.

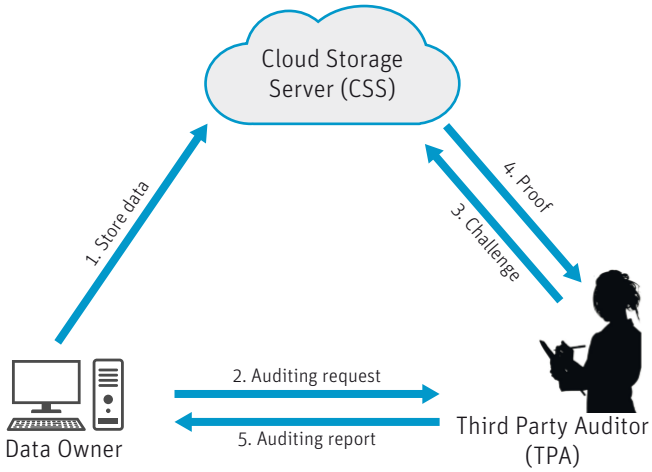


Fig. 2. Basic public auditing system.

B. Threat Models

A data owner assumes that a TPA is a reliable and honest entity that will verify the integrity of their data, but the TPA may be curious about that data. The TPA could thus be a threat for the data owner. To ensure the correct storage of the owner's data with the CSS, a privacy protection mechanism, which guarantees a TPA cannot access the owner's data, will thus be necessary.

A CSP also cannot be fully trusted; it can pose a threat to an owner's data. In order to save space, a CSP may remove data that is rarely accessed without any notification to the data owner. The outsourced data may be tampered with or even re-outsourced without notice by malicious a CSP. A CSP can also apply the wrong changes to an owner's data owing to system failure, management errors or other reasons, and hide these mistakes to protect their image. Undiscovered strangers may also be able to intrude on the cloud server and contaminate or erase an owner's data. When data is stored on a CSS and to respond to a TPA's query, CSS can use an authentic pair of data blocks as a substitute for the queried data blocks just to pass out the audit. The CSS can also retrieve the previously stored results of data that has been challenged simply to generate proof of data possession rather than to query the owner's data.

C. Our Method Goals

Motivated by data integrity and deduplication, we propose a new method for storage and auditing in cloud computing, based on the blockchain data structure. The proposed method achieves the following functions:

- **Confidentiality**: ensures the confidentiality of the owner's data against the TPA during the auditing process.
- **Batch auditing**: ensures that a mediator or TPA (depending on the auditing type) performs multiple auditing tasks, in a simultaneously way, received from different users.
- **Client-side deduplication** (storage efficiency): allows the mediator to eliminate duplicated files and file-blocks before sending the data to the cloud.
- **Private auditing**: allows only the mediator to verify the correctness of the data stored in the cloud.
- **Public auditing**: allows the TPA to check the correctness of the data stored in the cloud.
- **Data integrity**: ensures that the CSP cannot cheat and pass the auditing process without having stored the data intact.
- **Lightweight**: provides the model with low communication and computational overheads.

In this paper, we propose a new method that ensures both efficient storage based on data deduplication on the client side, and preserves data integrity auditing using blockchain technology in a cloud computing environment. The structure of this paper as follows: Section II outlines the various related works. Section III presents the different concepts used in our proposed method. Section IV provides a detailed description of our proposed method. Section V includes security analysis and performance evaluation. Finally, a conclusion is presented in Section VI.

II. Related Work

Many auditing protocols have been established to ensure the correctness of data stored in the cloud. Ateniese et al. [2] proposed a provable data possession (PDP) scheme. In this model, the third party auditor was allowed to statically check the correctness of the outsourced data without retrieving the data. The main goal of this model is to check that the server has the original data. Another improved version of this protocol is the E-PDP [3], which is 185 times faster compared to the first protocol. Proof of retrievability (POR) is another variation of PDP, proposed by Juels and Kaliski[4]. The main drawback of the above protocols is that they do not allow dynamic data auditing.

Erway et al. [5] proposed the concept of dynamic provable data possession (DPDP). According to this scheme, a data owner is allowed to modify the stored data. The main drawback of this scheme is that it cannot support public auditing. Wang et al. [6] resolved the above two problems by applying a Merkle Hash Tree (MHT) and presenting a public and dynamic auditing scheme, however, this scheme involves more computational costs during the updating and auditing phases.

Liu et al. [7] expanded MHT to rank-based MHT (R-MHT) with efficient verifiable fine-grained updates. Zhang and Blanton [8] improved the MHT scheme to include a balanced update tree. To minimize computation and communication costs, Zhu et al. [9] presented a new auditing scheme known as index-hash table-based public auditing (IHT-PA), however, it is inefficient for dynamic updating operations. Tian et al. [10] introduced a new scheme based on a dynamic hash table (DHT), which supports public and dynamic auditing. This scheme achieves better performance in the updating phases. Tang and Zhang [11] proposed a verifiable data possession (PVDP), which allows both private and public verifiability simultaneously, to check the integrity of

client files stored on a cloud server without downloading all those files. Li, Tan and Jia [12] proposed a simple efficient auditing scheme for checking the integrity of data stored in the cloud. This scheme supports dynamic operations and batch auditing.

Yu et al. [13] proposed an identity-based auditing scheme for checking the integrity of cloud data, but Xu et al. revealed that this scheme is vulnerable to data recovery attack. They thus presented a secure and efficient identity-based public auditing scheme using the RSA algorithm for cloud storage [14]. Lee et al. [15] presented a new data integrity check scheme for remotely acquired and stored stream data.

Yuan and Yu [16] proposed a public and constant cost storage integrity auditing scheme with secure deduplication (PCAD). Zhang et al. [17] introduced a fast asymmetric extremum content-defined chunking algorithm for data deduplication in backup storage systems. Gaetani et al. [18] proposed a blockchain-based database to ensure data integrity in cloud computing environments. El Khanboubi and Hanoune [19] proposed a new scheme exploiting blockchains to improve data upload and storage in the cloud.

III. Concepts Used in our Proposed Method

In our proposed method, we introduce the use of five concepts. Blockchain [20] is able to effectively ensure the integrity and authenticity of the exchanged data, and especially auditability by providing a private layer where cloud data is treated and stored in less time. The security of blockchain technology is enforced in a distributed and public way rather than relying on a central party to do so, as is the case for other databases. The blockchain has appeared as a fascinating technology which offers compelling and imperious properties about data integrity. A Merkle Hash Tree [21] is a binary tree that represents the data structure used in blockchain technology. A Third Party Auditor has the ability to check the data possession of the Cloud. The data deduplication technique eliminates redundant data and stores just one copy of each file or file-blocks (chunks) in order to minimize both network traffic and storage space. A mediator performs data deduplication to eliminate duplicated files and file-blocks.

A. Third-Party Auditor

A TPA, who has considerable computation and communication ability, is delegated by the cloud user to check the data possessed by the cloud. TPA is a semi-trusted entity with the expertise and ability to check the correctness of data on behalf of the data owner. The data owner, who employs the TPA to verify the integrity of their data, is alleviated from the burden of expensive auditing operations. Although the data owner has confidence in the TPA's data checking, they can be also a threat to the data owner. One of the most important issues in the data audit process is thus preventing data leakage and preserving the privacy of data.

B. Deduplication Technique

In this paper, we explore the technique of deduplication in the server where many thin clients are connected. For instance, many users from an enterprise-x may intend to outsource a large quantity of data to the cloud and many of these files or file-blocks are duplicates. It is therefore necessary to find and remove duplication within the data. Thus, we decided to use a mediator with the ability to manage client-side data deduplication. By transferring only a single copy of duplicate data, a deduplication system optimizes storage and bandwidth efficiency in cloud storage servers. Accordingly, client-side deduplication implies low communication and computation costs between the client and the CSP, and saves storage space. A file can be divided into many file-blocks (chunks) that can be part of many files. Chunking is an essential step achieving data deduplication, which permits a reduction in the storage space and alleviates the outgoing network traffic when uploading data to the cloud storage server.

Chunking is a challenging technique in the deduplication process, but it can be performed within several algorithms [22]: File-Level Chunking (FLC), Fixed-Size Chunking (FSC), Variable-Size Chunking (VSC), Content-Aware Chunking (CAC). In our proposal, we use a CAC algorithm, where the file is divided based on its content which improves the file-blocks reuse probability, unlike an FSC that splits a file into equally sized file-blocks which reduces the probability of using the same file-block in other files. Consequently, the CAC algorithm outperforms the FSC algorithm in terms of deduplication efficiency and has been extensively used in various storage systems.

C. Mediator

To reduce computational operations among users and to perform data deduplication using a central node in enterprise-x, we decided to use the concept of a mediator. A mediator manages the deduplication process internally in the server, so there is no security issue. The mediator has the ability to manage the storage of a user's data, and even to check the integrity of this data. The mediator has two tasks to perform:

- A client side deduplication to eliminate duplicated files and file-blocks before storing the data in the cloud. Accordingly, the quantity of stored data and the bandwidth used between the client and the Cloud server are both reduced.
- Check the integrity of a user's data stored in the cloud in the case of private auditing, where TPA is absent. It can be seen as an internal auditor, with the proviso that the mediator should have considerable expertise and ability to verify the correctness of the stored data.

IV. Description of the Proposed Method

Blockchain technology brings us many reliable and convenient services, such as preserving the integrity of data, however there are several security issues and challenges behind this innovative technique that should be overcome [23]. In our proposed method, each block in the blockchain database will only store the Merkle root, and the information of the file and the hash of the previous block. The files are not stored in the blockchain, rather they are stored in the CSP servers.

The mediator is trusted and allowed to see the content of the files and their hashes. It computes the file-blocks' hashes and the Merkle root, and then it compares them with a local database of Merkle roots and hashes stored in the previous operations in order to identify duplicated files/file-blocks. The TPA is semi-trusted and allowed to verify the integrity of the files, but it is prohibited from access to the content of the files. The CSP is semi-trusted and allowed to see the content of data, but it is obliged to follow the steps needed for the auditing process.

Each file gives rise to a Merkle hash tree. The Merkle hash tree allows a digest to be made of all the file-blocks linked to that block. File-blocks are not stored in the Merkle Tree, rather their hashes are stored in each node. If a small bit is changed in any file-block, there will certainly be a significant difference between the bit patterns of the resulting Merkle roots. Each Merkle root, generated from the hashes of file-blocks (leaf nodes) corresponding to a file, is stored in a new block in the blockchain with other information describing the file. The Merkle root is fundamental because it relies on the hashes of all underlying file-blocks. It therefore allows efficient and secure verification of data content.

In our proposed method, we use blockchain technology, where information for a file is stored in a block. Each block contains the user ID (U_{id}), the file ID (F_{id}), version number v, timestamp t, the number

of file-blocks N, the Merkle root n_0 and the hash of the previous block in the chain. One block B in the blockchain may correspond to the file F of the user U. The following block C may correspond to file L of the user J. However, the entire tree and the file-blocks, of a file, stored in the CSP database may correspond to one or more blocks in the chain, and thus to one or more users. The lengths in bytes of the different records in a block are: U_{id} 8, F_{id} 8, v 4, t 4, N 8, n_0 32, *HashPrev* 32. Fig. 3 displays the information stored in each block of the blockchain.

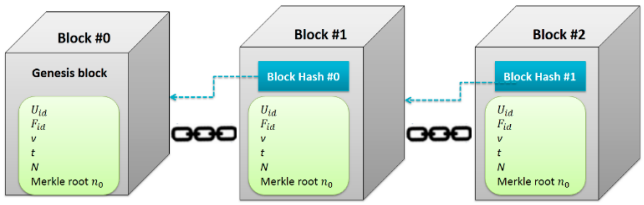


Fig. 3. Blockchain structure used in our proposal.

We decided to use MD5 hashing algorithm that generates a unique 32 chars string. Although the SHA-3 algorithms are more secure as compared to MD5, this latter is better in terms of speed and the hash string length produced is smaller than any other hashing algorithm. According to Yu Sasaki and Kazumaro Aoki [24], even though there have been many powerful collision attacks on MD5, the preimage resistance of MD5 has not been broken yet. Nevertheless, any other hashing algorithm may be applicable.

A. System Model for Private Auditing

In this mode of auditing, the mediator should have the ability to perform the auditing process.

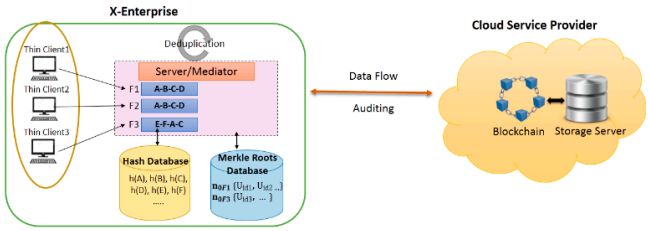


Fig. 4. System model for private auditing.

According to some statistics, more than 75% of the outsourced data in the cloud is not unique [25], and so the manipulation of deduplication could greatly reduce storage cost and the required space to store large data. The use of this technique in this system model thus ensures the maximum use of available storage space through the elimination of redundant data, and the amount of outsourced data and the bandwidth used between the enterprise and the CSP are also both reduced.

The main advantage of this system model is that the audit tasks are performed by an internal entity, which is the mediator, which implies low communication and computation costs. The mediator could therefore perform multiple auditing tasks simultaneously, and received from different users.

B. System Model for Public Auditing

In this mode of auditing, a third party auditor checks the integrity of the data stored in the cloud.

The auditing tasks are delegated to a competent external entity, the TPA, which implies more communication and computation costs. The auditing request could be sent directly by users to the TPA. We also use, in this system model, a technique of deduplication that reduces both storage cost and the bandwidth used between the enterprise and the CSP. A TPA could also perform multiple auditing tasks simultaneously, as received from different users.

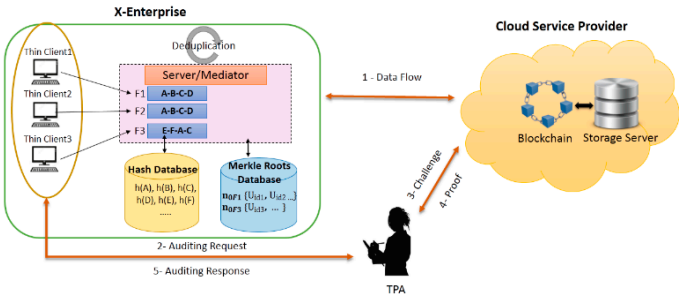


Fig. 5. System model for public auditing.

C. Storage Phase

As shown in our model (Fig. 4 and Fig. 5), when users intend to store their files on the cloud server, the mediator checks the existence of the entirety of each file or some of its file-blocks in the cloud storage server. The mediator therefore initiates data information: U_{id}, F_{id}, v, t, N corresponding to the file. Afterward, the mediator divides the file into N file-blocks using a CAC algorithm $(fb_0, fb_1, \ldots fb_{N-1})$ where $N=2^d$ and d is the depth of the Merkle tree, then he calculates its hashes $h(fb_i)$ with a secure hash function where $0 \leq i \leq N-1$, in order to compute the Merkle root n_0. Thereafter, he computes the Merkle root n_0 (*for* i *in* $N-2....0$: $n_i = h(n_{2i+1} \| n_{2i+2})$) of the file. After that, the mediator compares it with a local database of Merkle roots stored in the previous storage operations in order to identify duplicated files.

1. Case 1: The File Has Never Been Stored

If the generated Merkle root n_0 does not resemble any roots, the mediator keeps the root locally in the Merkle roots database with the U_{id} of the user in order to use them in the next storage operations. After that, the mediator performs file-blocks level deduplication by comparing the generated hashes with those located in the hash database (comparing the file-blocks' hashes will take less time than comparing file-blocks). There are two cases:

a) Case 1.1: Storing All File-Blocks

If the mediator did not find any identical file-block hashes, he stores the hashes of all file-blocks in the hash database, then outsources all file-blocks to the cloud storage server. A new block is then created in the blockchain, this block contains the file's information and the Merkle root that correspond to the file. After that, the CSP stores the entire tree with the file-blocks corresponding to that file. The concerned user thereafter maintains a pointer to the block that corresponds to his file. So that, each user preserves a file ID list; this list contains pointers, each pointer points directly to the particular block in the blockchain corresponding to that file.

b) Case 1.2: Storing Some File-Blocks

If the mediator finds some identical hashes, he ignores them and stores the other file-blocks hashes (unduplicated hashes) in the hash database. After that, he stores the unduplicated file-blocks in the cloud storage server. A new block is then created in the blockchain, this block contains the file's information and the Merkle root that correspond to the file. After that, the CSP stores the entire tree with the uploaded file-blocks corresponding to that file. Hereafter, the concerned user maintains a pointer to the block that corresponds to his file. In this case, the mediator does not need to upload all the file-blocks because some of them have been stored previously by the same or other users, so, the mediator ignores the duplicated file-blocks, which reduces disk utilization.

2. Case 2: The File Has Already Been Stored

a) Case 2.1: The Same User Tries to Store the Same File

If the generated Merkle root n_0 resembles a root which has already been registered in the Merkle roots database, and if the current U_{id} already exists in the user's ID list related to this Merkle root, the mediator informs the file owner that he has already stored this file in a previous storage operation. In this case, the mediator does not need to send the file to the CSP to be stored, which reduces the disk's utilization.

b) Case 2.2: Another User Tries to Store the Same File

If the generated Merkle root n_0 resembles a root which has already been registered in the Merkle roots database, and if the current U_{id} does not exist in the users ID list related to this root, the mediator adds the new U_{id} to the user's ID list corresponding to this Merkle root, and then stores n_0 in a new block in the blockchain with the file information. In this case, the mediator does not need to send the file to the CSP to be stored, because it has been stored by another user in a previous storage operation. Finally, the user maintains a pointer to the block that corresponds to his file.

D. Execution Flow for Storage and Auditing Files

The execution flow for the storage and auditing of files is shown in the Fig. 6.

E. Auditing Phase

For the auditing process, we follow the technique of verification used in [26]. So to verify the integrity of a file:

- The mediator/TPA (depending on auditing type) computes the generator seed $r = h^P(n_0)$ where leaves are divided into P chunks.
- After that, the mediator/TPA derives leaf numbers in each P chunk as: for j in 0 ... P – 1: $l_j = G(r, j)$ with G some cryptographic pseudo-random number generator (PRNG).
- Then, the mediator/TPA sends the leaf numbers $\{lj\}$ to the CSP.
- The CSP provides the appropriate sibling information to the mediator/TPA, which allows the mediator/TPA to compute the new Merkle root n'_0.
- The mediator/TPA verifies whether $n_0 = n'_0$ or not.
- The mediator/TPA then calculates the new generator seed $r' = h^P(n'_0)$.
- The mediator/TPA deduces the leaf numbers $l'_j = G(r', j)$.
- Hereupon, the mediator/TPA checks whether $l'_j = l_j$ for each j *in* $0 \ldots P - 1$, and if they match, then the file is verified.
- Finally, the mediator/TPA informs the cloud client of the results.

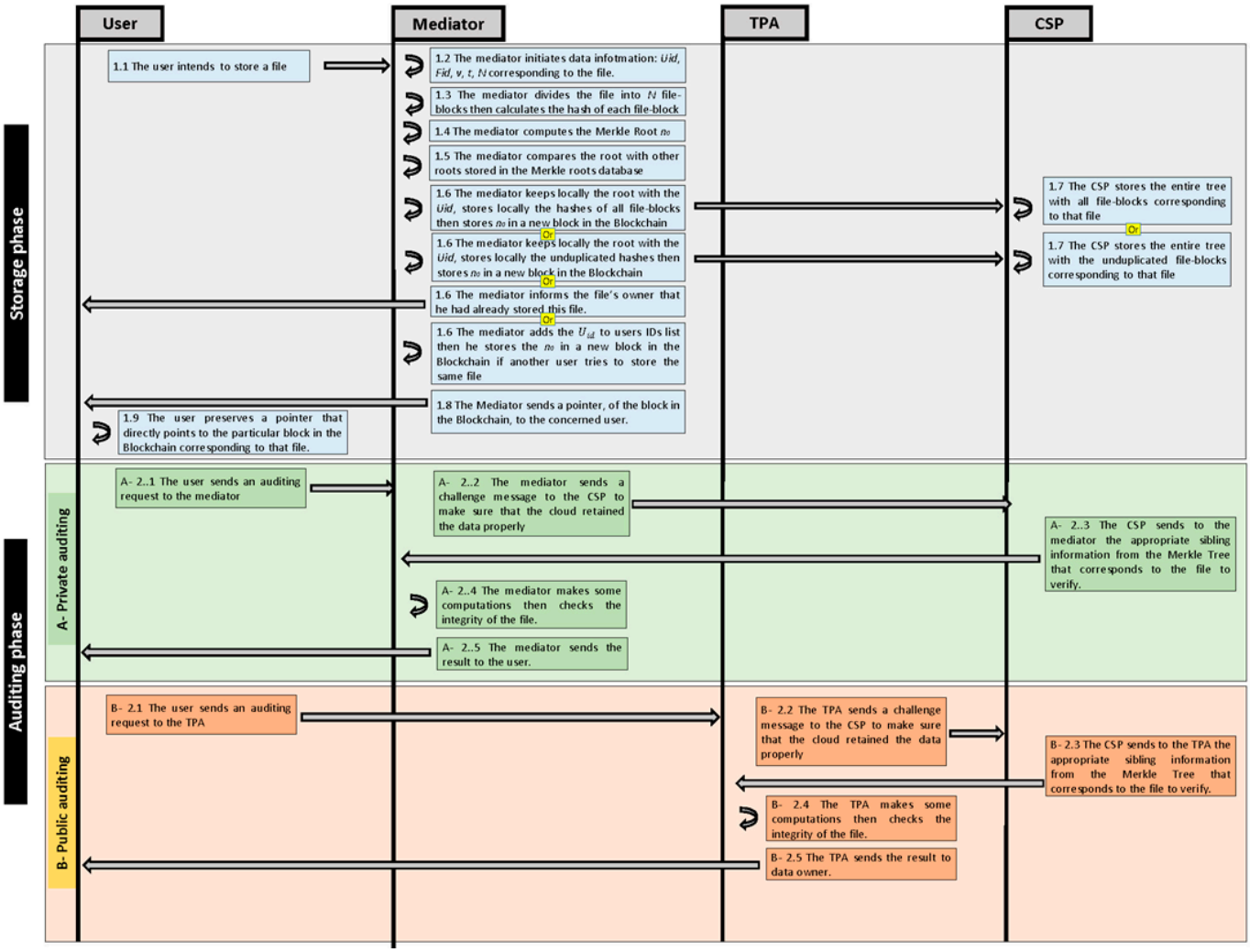


Fig. 6. Execution flow for storage and auditing of files.

V. Security Analysis and Performance Evaluation

Several data integrity auditing schemes have implemented data deduplication in the cloud server side. This way of working has high computational costs. Other schemes follow a fixed-size chunks method which is simple and extremely fast, but this approach suffers from low deduplication efficiency. In our method, deduplication is performed on the client side by the mediator and using a content-aware chunking algorithm. Instead of saving three or four copies of the same file/chunk in the cloud, deduplication allows the elimination of all the redundant data and stores only one copy of the file/chunk that belongs to one or multiple cloud users. This technique decreases communication, computation and storage costs.

Note that in our proposal, we indicate the use of two types of auditing, private auditing and public auditing, to verify the correctness of the data stored in the cloud. It depends on the need of the enterprise; if the mediator has the ability to verify the accuracy of the data, in this case, he manipulates a private auditing, or he prefers the use of an external auditor to verify their data (public auditing).

Owing to the technique that we used during the auditing process, the TPA will have no idea about the owner's data, which implies that the confidentiality of the data is ensured against auditors.

The mediator or TPA (depending on the auditing type) could perform multiple auditing tasks simultaneously, received from different users. In a case where the mediator/TPA receives several auditing requests for the same file from different users, it may be ineffective to handle them as individual tasks rather than to batch them together and perform only one audit task by interacting with the CSP to check the data integrity. After that, it replies all concerned users by the auditing result. The deduplication technique is thus not only efficient for data storage because it reduces storage cost and the required space to store a large data, but it is also efficient for the auditing process where multiple users want to verify the same file while reducing the communication and computation cost between auditor and CSP.

The use of blockchain in our proposal is mainly applicable in scenarios where the data history is very important. This method is practical for real application scenarios, such as in the justice domain where judgments must be stored and must not be modified. It can also be useful for real estate agents, to register property titles for example, where it is forbidden to modify this type of data. It could be also used for storing medical records or even collecting taxes.

To demonstrate that our proposed method is efficient, we performed experiments through an application developed using Java and PHP languages, on a computer with an Ubuntu 17 OS running on an Intel CoreTM i3 CPU with a 2.27 GHz clock and 4 GB RAM. The remote storage was implemented using MySQL. The size of each file is increased by 200 MB. We used large files to show the usefulness of thededuplication technique. Fig. 7 shows a plot of computation time in seconds against input size in MB.

We can see that the computation time for an unduplicated file is greater than the computation time for a duplicated file, which shows that the computation time and the bandwidth, used between the enterprise and the CSP, are both reduced thanks to the deduplication technique.

Several systems perform a public data integrity auditing where the responsibility of data integrity verification is delegated to a third party auditor as in [2]-[4],[6]-[10],[14],[15]. Other schemes manipulate a private auditing where the data owner checks the integrity of the externalized data as in [5]. However, in some cases, it is not practically feasible for the data owner to verify the data integrity all the time. Hence, our approach supports both public and private auditability property.

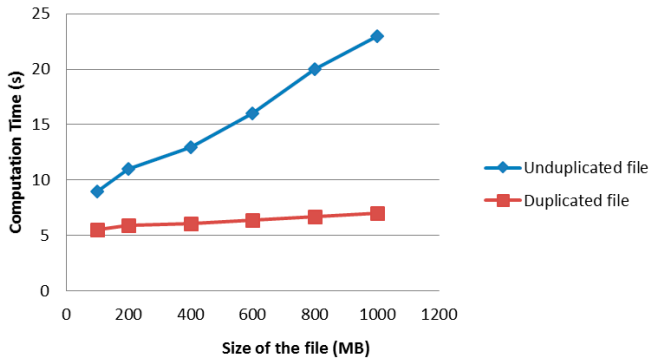


Fig. 7. Comparison of the time computed in storage operations for duplicated and unduplicated files of different sizes.

Some researches mention the use of data deduplication technique to improve storage systems as in [16], [17]. Besides, the decentralization and security characteristics of blockchain technology have attracted many researchers to propose various schemes exploiting blockchain-based databases to ensure data integrity and improve data storage in cloud computing environments as in [18], [19]. The value-added of our method over these methods is that our proposal permits both data deduplication, which guarantees the storage efficiency, and data integrity auditing that verifies the correctness of the outsourced data.

In our system, the deduplication is performed at the file level or block level on the client side. Hashes of files or blocks are computed and stored, and if the hash for any new file or block is found to be present in the stored hashes, then the copy is removed, which permits to eliminate duplicated files and file-blocks before storing the data in the cloud. Accordingly, the quantity of stored data and the bandwidth used between the client and the Cloud server are both reduced. The main benefit of manipulating deduplication in our system is that storage efficiency is increased and network efficiency is enhanced.

As with any system, it is necessary to have a fault tolerance property that enables the system to continue its operation properly in the event of the failure of any of its components.

VI. Conclusions

In this paper, we demonstrated the feasibility of using deduplication and blockchain technologies to manage storage and auditing processes for cloud data.

We proposed the concept of the mediator, who performs client-side data deduplication to identify and remove duplicated files and file-blocks using a content-aware chunking algorithm, which reduces the computation costs on the user's side, communication costs in the channel and storage costs in CSP.

Our method relies on blockchain to perform the auditing process in a transparent and lightweight way. The main goal of using blockchain is that TPA can check the integrity of the outsourced data without gaining any knowledge of the user's data. Consequently, we can consider the use of blockchain as a new model for trust.

References

[1] P. Mell and T. Grance, "The NIST Definition of Cloud Computing Recommendations of the National Institute of Standards and Technology," *Nist Spec. Publ.*, vol. 145, p. 7, 2011. https://doi.org/10.1136/emj.2010.096966.

[2] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson *et al.*, "Provable data possession at untrusted stores," *Proc. 14th ACM Conf. Comput. Commun. Secur. - CCS '07*, p. 598, 2007. https://doi.org/10.1145/1315245.1315318.

[3] G. Ateniese, R. Burns, R. Curtmola, J. Herring, O. Khan, L. Kissner *et al.*, "Remote data checking using provable data possession," *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 1, pp. 1–34, 2011. https://doi.org/10.1145/1952982.1952994.

[4] A. . Juels and B. S. Kaliski Jr., "Pors: Proofs of retrievability for large files," *Proc. ACM Conf. Comput. Commun. Secur.*, pp. 584–597, 2007. https://doi.org/10.1145/1315245.1315317.

[5] C. C. Erway, A. Küpçü, C. Papamanthou, and R. Tamassia, "Dynamic Provable Data Possession," *ACM Trans. Inf. Syst. Secur.*, vol. 17, no. 4, pp. 1–29, 2015. https://doi.org/10.1145/2699909.

[6] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, "Enabling public auditability and data dynamics for storage security in cloud computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 5, pp. 847–859, 2011. https://doi.org/10.1109/TPDS.2010.183.

[7] L. Chang, J. Chen, L. T. Yang, X. Zhang, C. Yang, R. Ranjan *et al.*, "Authorized public auditing of dynamic big data storage on cloud with efficient verifiable fine-grained updates," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 9, pp. 2234–2244, 2014. https://doi.org/10.1109/TPDS.2013.191.

[8] Y. Zhang and M. Blanton, "Efficient Dynamic Provable Possession of Remote Data via Update Trees," *ACM Trans. Storage*, vol. 12, no. 2, pp. 1–45, 2016. https://doi.org/10.1145/2747877.

[9] Y. Zhu, G. J. Ahn, H. Hu, S. S. Yau, H. G. An, and C. J. Hu, "Dynamic audit services for outsourced storages in clouds," *IEEE Trans. Serv. Comput.*, vol. 6, no. 2, pp. 227–238, 2013. https://doi.org/10.1109/TSC.2011.51.

[10] T. Hui, Y. Chen, C. Chen Chang, H. Jiang, Y. Huang, Y. Chen *et al.*, "Dynamic-Hash-Table Based Public Auditing for Secure Cloud Storage," *IEEE Trans. Serv. Comput.*, vol. 10, no. 5, pp. 701–714, 2017. https://doi.org/10.1109/TSC.2015.2512589.

[11] C. ming Tang and X. jun Zhang, "A new publicly verifiable data possession on remote storage," *J. Supercomput.*, vol. 75, pp. 77–91, 2019. https://doi.org/10.1007/s11227-015-1556-z.

[12] A. Li, S. Tan, and Y. Jia, "A method for achieving provable data integrity in cloud computing," *J. Supercomput.*, vol. 75, pp. 92–108, 2019. https://doi.org/10.1007/s11227-015-1598-2.

[13] Y. Yong, L. Xue, M. H. Au, W. Susilo, J. Ni, Y. Zhang *et al.*, "Cloud data integrity checking with an identity-based auditing mechanism from RSA," *Futur. Gener. Comput. Syst.*, vol. 62, pp. 85–91, 2016. https://doi.org/10.1016/j.future.2016.02.003.

[14] Z. Xu, L. Wu, M. K. Khan, K. K. R. Choo, and D. He, "A secure and efficient public auditing scheme using RSA algorithm for cloud storage," *J. Supercomput.*, vol. 73, no. 12, pp. 5285–5309, 2017. https://doi.org/10.1007/s11227-017-2085-8.

[15] K. M. Lee, K. M. Lee, and S. H. Lee, "Remote data integrity check for remotely acquired and stored stream data," *J. Supercomput.*, vol. 74, no. 3, pp. 1182–1201, 2018. https://doi.org/10.1007/s11227-017-2117-4.

[16] J. Yuan and S. Yu, "Secure and constant cost public cloud storage auditing with deduplication," in *2013 IEEE Conference on Communications and Network Security, CNS 2013*, 2013, pp. 145–153. https://doi.org/10.1109/CNS.2013.6682702.

[17] Z. Yucheng, D. Feng, H. Jiang, W. Xia, M. Fu, F. Huang *et al.*, "A Fast Asymmetric Extremum Content Defined Chunking Algorithm for Data Deduplication in Backup Storage Systems," *IEEE Trans. Comput.*, vol. 66, no. 2, pp. 199–211, 2017. https://doi.org/10.1109/TC.2016.2595565.

[18] E. Gaetani, L. Aniello, R. Baldoni, F. Lombardi, A. Margheri, and V. Sassone, "Blockchain-based database to ensure data integrity in cloud computing environments," *CEUR Workshop Proc.*, vol. 1816, pp. 146–155, 2017.

[19] Y. El Khanboubi and M. Hanoune, "Exploiting Blockchains to improve Data Upload and Storage in the Cloud," vol. 11, no. 3, pp. 357–364, 2019.

[20] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," *Www. Bitcoin.Org*, p. 9, 2008. https://doi.org/10.1007/s10838-008-9062-0.

[21] R. C. Merkle, "I NFORMAT I ON SYSTEMS LABORATORY By," 1979.

[22] Siva Sankar K. Venish A., "Study of Chunking Algorithm in Data Deduplication," *Proc. Int. Conf. Soft Comput. Syst. Adv. Intell. Syst. Comput.*, vol. 398, pp. 319–329, 2016. https://doi.org/10.1007/978-81-322-2674-1.

[23] I. C. Lin and T. C. Liao, "A survey of blockchain security issues and challenges," *Int. J. Netw. Secur.*, vol. 19, no. 5, pp. 653–659, 2017. https://doi.org/10.6633/IJNS.201709.19(5).01.

[24] Sasaki Y., Aoki K. (2009) "Finding Preimages in Full MD5 Faster Than Exhaustive Search" *In: Joux A. (eds) Advances in Cryptology - EUROCRYPT 2009. Lecture Notes in Computer Science,* vol 5479. Springer, Berlin, Heidelberg.

[25] J. Gantz and D. Reinsel, "The Digital Universe Decade – Are You Ready?," *Idc*, vol. 2009, no. May, p. 16, 2010.

[26] F. Coelho, "An (almost) constant-effort solution-verification proof-of-work protocol based on Merkle trees," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5023 LNCS, pp. 80–93, 2008. https://doi.org/10.1007/978-3-540-68164-9_6.

Mohamed El Ghazouani

He is a Doctor at the Computer Science Department of Cadi Ayyad University in Marrakesh, Morocco. He received his Master's degree in Information Systems Engineering from the same university in 2015. His research interests are Computer Science, Cloud Computing Security, Big Data and Blockchains.

My Ahmed El Kiram

He is a full Professor of Computer Science at the Department of Computer Science, Faculty of Science Semlalia, Cadi Ayyad University, Morocco. He writes and presents widely on issues of IT Security, Cryptographic Systems and Cloud Computing. He is the author of numerous publications related to his research interests.

Latifa Er-Rajy

She is a Doctor at the Computer Science Department of Cadi Ayyad University in Marrakesh, Morocco. Her research interests are Android Applications, Mobile security and Security in Cloud Computing. She is the author of several publications related to his research interests.

Yassine El Khanboubi

Hereceived his Master's degree in Information Systems Engineering from Cadi Ayyad University Marrakesh, Morocco. He is currently a PhD candidate at the Hassan II university. His research interests include Computer Network Security, Mobile and Wireless Communication Security.

Tracking News Stories Using Blockchain to Guarantee their Traceability and Information Analysis

Francisco Jurado*, Oscar Delgado, Álvaro Ortigosa

Department of Computer Engineering, Universidad Autónoma de Madrid, Madrid (Spain)

Received 15 nov 2019 | Accepted 24 Feb 2020 | Published 25 June 2020

Abstract

Nowadays, having a mechanism to guarantee the traceability of the information and to monitor the evolution of the news from its origin, and having elements to know the reputation and credibility of the media, analyze the news as well as its evolution and possible manipulation, etc. is becoming increasingly significant. Transparency in journalism is currently a key element in performing serious and rigorous journalism. End-users and fact-checking agencies need to be able to check and verify the information published in different media. This transparency principle enables the tracking of news stories and allows direct access to the source of essential content to contrast the information it contains and to know whether it has been manipulated. Additionally, the traceability of news constitutes another instrument in the fight against the lack of credibility, the manipulation of information, misinformation campaigns and the propagation of fake news. This article aims to show how to use Blockchain to facilitate the tracking and traceability of news so that it can provide support to the automatic indexing and extraction of relevant information from newspaper articles to facilitate the monitoring of the news story and allows users to verify the veracity of what they are reading.

Keywords

Blockchain, Smart Contract, Traceability, News Stories, Journalistic Transparency.

DOI: 10.9781/ijimai.2020.06.003

I. Introduction

NOWADAYS, the news media are subject to the social scrutiny because of the lack of credibility, the manipulative media, the misinformation campaigns, and the propagation of fake news [1].

In [2], Lazer *et al.* point out that the main difference between fake news and true news relies on lack editorial norms and processes that ensure the accuracy and credibility of the information. Thus, to arrange a way that allows guaranteeing these editorial processes (or at least part of them) can suppose a big step in the fight against the above-mentioned issues. Also, in their work, Lazer *et al.* mentioned as an effective intervention to empowering individuals to evaluate the news. This way, end-users and fact-checking agencies like factcheck.org, snopes.com, politifact.com, hoax-slayer.com, truthorfiction.com, urbanlegends.about.com, newtral.es, or maldita.es, need a way to check and verify the information published in different media.

Is in this scenario where the *journalistic transparency* [3]–[8] has emerged as an essential concept to face the aforementioned issues. Applying the principle of journalistic transparency allows the readers to corroborate the information by directly follow the editorial processes and easily going to the source of the news and checking their veracity. So much so that transparency in journalism is currently a key element in performing serious and rigorous journalism. This transparency principle allows direct access to the source of essential content to contrast the information it contains and even to know whether it has been manipulated. Giving the right tools to guarantee the journalism process and allowing the users to verify the information are key concepts because humans tend to spread false news more than truth [9].

Additionally, allowing to check the journalism processes followed in the news stories, to track the evolution of the news reports and the relevant data and information it contains as they change over time, and therefore to trace how the related news evolves, constitute other instruments to face the previous issues. This is not only useful for end-readers but for fact-checking agencies and those tools that perform automatic indexing and extraction of relevant information of news. Once the information has been verified and fact-checked, these tools need a way to guarantee that the extracted data have not changed.

Combining these key matters in just one approach will suppose a starting point to solve the issues we mentioned at the beginning of the section.

However, as far as we know, there is no mechanism that allows guaranteeing to check the transparency and traceability of the information by monitoring the evolution of the news from the event that causes them, its possible manipulation during the dissemination process, and estimating the reputation and credibility of the media.

In this scenario, this article aims to detail the use of blockchain to facilitate the principle of journalistic transparency by allowing the tracking and traceability of news stories. Blockchain will be used as a tool to guarantee the traceability process and information analysis. This way, we will be able to support the automatic indexing and extraction of relevant information from newspaper articles to facilitate the monitoring of the news story and allows end-users readers and fact-checking agencies to verify the truthfulness of what they are reading.

The rest of the article is structured as follows: in section II we introduce the background and basic concepts; section III details the desired requirements we established for the proposal; section IV defines

* Corresponding author.
E-mail addresses: francisco.jurado@uam.es

the key elements of the proposal; section V details the operations we can perform; VI focuses on the real cost of using the approach; and finally, section VII provides some conclusions and future works.

II. Background

In this section, we provide the basis of our proposal, that is, the current situation about the application of the journalism transparency, and introduce blockchain and digital signature to guarantee the data integrity and authorship requirements we need to achieve the traceability of news.

A. Journalism Transparency

Journalism transparency is an ethical principle emerged as a solution to improve accountability and journalists' credibility [3], [8]. Following this principle, quality journalism should reveal how information was obtained, and thus, the readers and fact-checking agencies can go to its origins and check it [8], [10], [11].

In this context, some initiatives have raised in order to create guidelines and standards. Thus, the Journalism Trust Initiative (https://jti-rsf.org) appeared with the aim to create an agreed set of trust and transparency standards.

Similarly, The Trust Project (https://thetrustproject.org/) is a consortium of news companies that appeared with the aim of developing transparency standards to easily assess the quality and credibility of journalism and other standards for fairness and accuracy, a journalist's background, and the work behind a news story. Among their initiatives, The Trust Project defined what they call the Trust Indicators, a list of standardized disclosures about the news organization's ethics.

In line with this idea, the Newsroom Transparency Tracker (https://www.newsroomtransparencytracker.com/) is a tool that helps to determine the trustworthy of a news agency, by displaying the kind of public information available on a wide range of journalistic policies and practices.

Also using transparency indicator, we can mention the Transparent Journalism Tool (TJ-Tool) developed by the Spanish online journal Público (https://publico.es). This tool generates what the journal calls a transparency map. In essence, for each news item, it displays the author of the news, the dates of creation and edition, the list of reference documents, the people mentioned in the information, etc.

For its part, opting for the use of a social approach for fact-checking, WikiTribune [12] uses a collaborative approximation to write evidence-based news articles, where journalism professionals and volunteers collaborate together.

All in all, the main journals and news agencies claim to adhere to ethical principles and/or follow good practices and trust standards. Despite that, as far as we know there is no mechanism to ensure their fulfilment to guarantee journalism transparency, and this is where our approach of using blockchain comes into play.

B. Blockchain as the Key to Guaranteeing the Data Integrity Within the Chain

Essentially, a blockchain is a simple concept: a distributed and secure data registry, which guarantees the integrity of the stored information.

Despite its enormous potential, the *blockchain* concept has a modest and recent origin. As defined today, it was first described as an auxiliary technology of Bitcoin in 2009 [13], where it is used as a secure mechanism to store economic transactions between participants. Its recent explosion in popularity is due to the possibility of storing any type of digital data, guaranteeing its integrity.

This automatically enables many new possible uses for the technology: certification of documentation as mortgages, securities or any other official document [14], [15]; assets or intelligent objects [16], [17], which can make decisions based on the information stored in the blockchain; distributed security market, deposit and custody services [18], which would resolve disputes between customers and merchants; voting systems [19], [20]; or improvements in the supply chain for all types of products [21]–[23].

The blockchain technology provides some desirable characteristics, namely: **immutability**, **accountability**, and **availability** and **universal access**. These characteristics automatically raise the level of security, transaction verifiability, operational transparency and privacy of the secured information:

- *Security*: The decentralized nature of blockchains could guarantee that data remains *available* even in the case of failure of a substantial number of nodes. Due to its intrinsic immutability, a blockchain also assures the *integrity* of the data once it is recorded in it.
- *Transaction verifiability*: In a blockchain, any participant can validate transactions by itself, without relying on a centralized judge. Usually, the roles of the nodes are also distributed, in such a way that the centralization of interests is discouraged.
- *Transparency*: Usually, all participants in a blockchain share the same data and operations, whose security is guaranteed by a distributed consensus algorithm. This provides an accurate and consistent database for all participants, although their permission of access to information can be changed in some blockchains configurations.

In summary, the autonomous verifiability of transactions without the possibility of tampering or the necessity of a third-trusted party is especially useful in scenarios where the parties involved have conflicting interests. In this case, none of the parties can be the owner of the data, to avoid the possibility of manipulation.

C. Digital Signature to Guarantee the Authorship

As stated before, a typical blockchain guarantees the integrity of the stored data, but not its authenticity. This property must be provided with external cryptographic primitives, like *digital signatures*. A digital signature is a set of data associated with an electronic document and whose basic functions are:

- Identify the signer unequivocally.
- Ensure the integrity of the signed document, that is, guarantee that the signed document is exactly the same as the original and has not been altered.
- Ensure the *non-repudiation* of the signed document. This is a usually forgotten property, but very important in the proposed scheme.

To achieve this, each participant must create a pair of public/private keys, typically by using an x509 certificate. This certificate must be issued by a trusted Certificate Authority (CA).

III. Desired Requirements for the Proposal

This section details the functional and non-functional requirements we set while designing our proposal to achieve our goal.

A. Functional Requirements

In this article, we are aimed to address the issues about the lack of credibility, the manipulative media, the misinformation campaigns, and the propagation of fake news discussed in the previous sections, by enhancing the journalism transparency using blockchain.

To do so, we propose an architecture that will allow tracking the news stories and guaranteeing their traceability and information analysis.

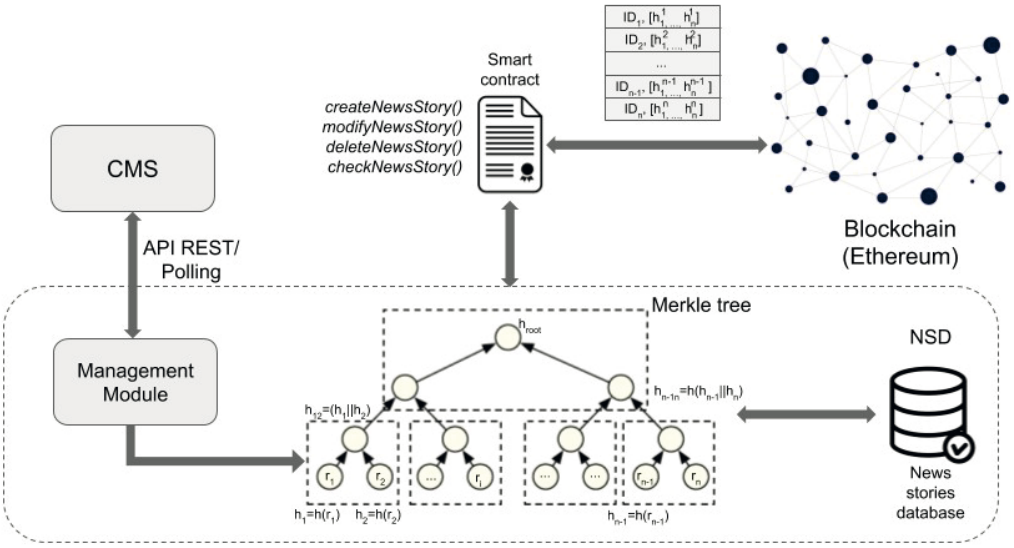


Fig. 1. Scheme of the proposal.

Thus, we established the next list of requirements for our approach:

- *Direct access to the source of information*: This requirement will easy the application of the journalistic transparency principle, what is essential to contrast the information contained in the news and to know whether it has been manipulated. This requirement is closely linked with the *transparency* characteristic of the blockchain.
- *Tracking changes of news stories*: This requirement will allow the traceability of the news, allowing analyze their evolution as they evolve over time. This requirement is related to the *transaction verifiability* in the blockchain.
- *Data and metadata integrity*: This requirement will be essential to guarantee the traceability process and information analysis. It is connected to the immutability characteristic in the blockchain.
- *Authorship insurement*: This requirement will provide a way to guarantee the authorship of the news, that is, the journalist or news agency who provides the information. This will be the key point where we will introduce the digital signature.
- *Date insurement*: This requirement will allow guaranteeing when the news was written and/or modified and, together with the previous requirement, by who. This is associated with the transaction verifiability in blockchain and the digital signature.

By achieving these requirements, we have ensured the journalism process, but also we have created a trust mechanism we have called the *fact-checking propagation*. That is, once a claim has been fact-checked, the linked data in the blockchain and the smart-contracts will guarantee and easily propagate the evidence gathered throughout the graph.

B. Non-functional Requirements

The main non-functional requirements of the proposal are:

- *Simplicity*: Keep the architecture as simple as possible, and with the lowest possible degree of coupling. This way the modifications to the existing systems are kept to a minimum.
- *Low computational requirements*: Keep the execution costs and processing time as low as possible, due to the huge potential number of data to be processed. To achieve this, our proposal will use an intermediate data structure called Merkle tree, in order to avoid the scalability problems of public blockchains.
- *Easy implementation*: Do not use complex smart contracts, which facilitates development and reduces execution costs. Smart contracts do not implement news treatment "logic"', but only the minimal functions needed to manage the storage of news stories (creation, modification, etc.)

This way, our proposal remains flexible and modular, by using a public blockchain as a secure storage mechanism to guarantee the integrity of the protected news stories.

IV. Architecture Key Elements

In this section, we describe the key elements of the proposed scheme to use public blockchains in the tracking of news. The rest of the subsections describes and analyzes the architecture and operation mode of our proposal, including its basic elements, namely:

- Management module
- Merkle trees
- News stories database
- Smart contract and blockchain

All these components are discussed in detail below, and a full scheme can be found in Fig. 1 in order to allow the reader a better understanding of the approach.

A. Management Module (MM)

This constitutes the central piece and the one in charge of managing all operations. It is also the entry point to the system, for which it will expose an API REST for integration with external tools. This API supports the basic CRUD operations (Creation, Read, Update, Delete) on each news story.

Thus, when the Content Management System (CMS) of news agencies or newspaper publishers creates a news story, the MM will support two modes of operation:

- *Active*: Active: the MM performs a periodically polling to the CMS to retrieve the new stories created during the last period.
- *Passive*: the CMS includes a small gateway module, which is activated when a new story is created and makes an appropriate call to the MM's REST API.

In any case, once the MM detects the creation of a news story, the certification process begins. This process is detailed in a later section.

B. Merkle Trees

Due to the transaction and storage costs and even to processing

capabilities, the protection of each individual news story with a public blockchain can be infeasible without an intermediate data structure.

For this reason, our architecture will make use of a data structure known as *Merkle tree* [24]. This construction is widely used in cryptography and computer science problems such as database integrity verification [25], peer-to-peer networks [26] and, of course, blockchains is not an exception [27].

A Merkle tree is a binary tree data structure in which every node contains the cryptographic hash of the concatenation of its child nodes contents. Due to this recursive way of constructing itself, the tree root contains statistical information of the rest of nodes, and the modification of any node content will cause the complete change of the value of the root. This way, the integrity of an arbitrary amount of data can be efficiently assured by arranging it in a Merkle tree form and securely storing the contents of its root node.

For every news story, our architecture maintains a Merkle tree storing the cryptographic hash of every necessary element in a node, namely: news author, publication/modification/deletion date, title, subtitle, body content, main picture, related documents (such as interviews, gazette publications, press conferences), etc.

To assure the root node of the Merkel tree we need what is called a *smart contract* (see section III.F). Therefore, when a new story is created, or an existing one is modified or deleted, the tree is re-calculated and the news root is updated in the blockchain.

As the hash function, the proposal can make use of any cryptographic hash function, such as the SHA3 family, which can produce outputs from 224 to 512 bits in length. However, in this work, we consider hashes of 256 bits per new story, which present a good balance between security and cost of storage.

C. News Stories Database (NSD)

As shown in the previous section, a Merkle tree only stores the hashes of the content, not the content itself. For this reason, the architecture includes the possibility of including an optional *News stories database* (NSD), which stores all the resources and elements of a news story. This way, the verification process, which will be described below, can be achieved more easily.

It is important to note that our proposal does not impose any restriction over this NSD. Typically, it could be a traditional relational database, internal or external to the organization.

D. Smart Contract and Blockchain

Finally, probably the most important element of the system is the *smart contract*, implemented in the Ethereum blockchain. This is the element that actually stores and secures the cryptographic proofs contained in the corresponding Merkle tree and implements the operations we can perform on them.

To demonstrate the viability and real performance of our architecture, the proposed smart contract has been implemented in Solidity language (see ANNEXE A. for the whole code) and deployed to the Ethereum Ropsten testnet at the address 0x8f737f448de451db9b1c046be7df3b48839673a1, where can be verified with any blockchain explorer like Etherscan.io. It is important to note that this is a basic contract, and that *should be used only for academic or educational purposes*.

The implemented operations in the smart contract are described in Table I. Of course, all these operations are authenticated, and can only be called by the owner of the architecture.

V. Operations of the Proposal

Once all the key elements have been analysed, this section describes the operation mode of our proposal.

A. Data to Work With

The core element is the concept of the *news story*. For each news story, journalists will provide its corresponding author, publication/modification/deletion date, title, subtitle, body content, main picture, and related content (such as interviews, gazette publications, press conferences) in any kind of format (text, audio, video), etc.

In our architecture, a news story (NS) is just an unordered set of resources **r** of length **n**, so that NS = (r1,..., rn). Of course, this set can be formed by any number and type of resources (typically text, images, videos, etc.).

From here, we set a protocol divided into two basic stages: **certification** and **verification**. The former describes the process that generates a cryptographic proof for each news story, which will allow us to verify its integrity at any time. The *verification* stage, on the other hand, allows any independent party to verify that the currently published news story matches exactly the initially published news and, if not, how, when and by whom it was changed. Both stages are described in detail below.

B. Certification

Essentially, the news stories are processed as follows (see Fig. 2):

1. A new story is created, processed and inserted in the CMS or similar software used by the journalist.
2. Our architecture periodically pulls this CMS and retrieves the news stories.
3. For each news story **NS**, a new Merkle tree $\mathbf{T}_{NS}$ is created, including

TABLE I. Operations Defined In The Smart Contract

Operation	Input arguments	Output	Description
createNewsStory()	NS_{ID}, h_{root}	-	This function is called to secure a news story with ID NS_{ID} and hash root h_{root} in the blockchain. If there exists already an entry with that identifier, it exits with no action.
modifyNewsStory()	NS_{ID}, h_{old}, h_{new}	*True* is modification is successful *False* if news story to modify does not exist	This function is called when a news story has been modified in some way since its inclusion in the system. It checks if the new story with ID NS_{ID} exists in the system and, if so, it changes its value to h_{new} and returns *True*. Otherwise, it returns *False*.
deleteNewsStory()	NS_{ID}	*True* is deletion is successful, *False* otherwise	This function deletes an existing news story with ID NSID. Due to the intrinsic immutable nature of public blockchains, data is not actually deleted, but marked as removed in the smart contract.
checkNewsStory()	NS_{ID}, h_{check}	*True* and timestamp of the certification if there exists a new story with ID NS_{ID} and hash h_{check}. *False* otherwise	This function is used during the verification phase, to check whether a news story was published as it appears today or has been modified.

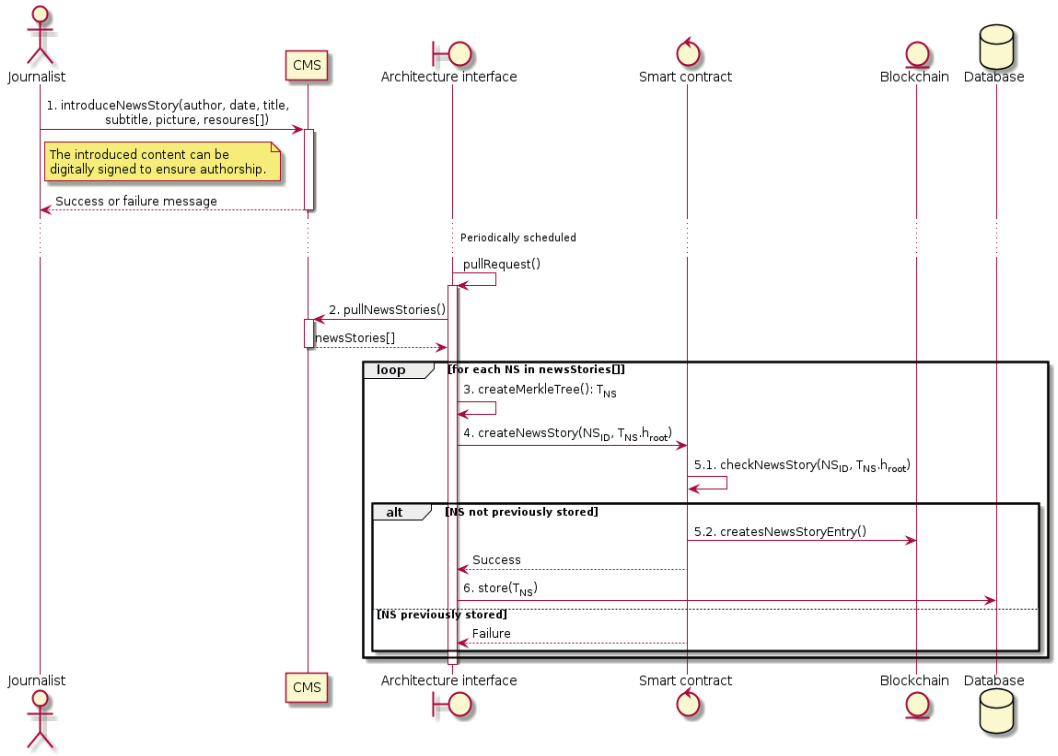


Fig. 2. Sequence diagram to perform the certification process while journalists introduce news stories in the CMS

each element ri of the NS as a leaf. The inclusion order items are arbitrary, but it has to be respected in the verification stage.

4. Once the construction of the $\mathbf{T}_{NS}$ is finished, its root leaf hash value, $\mathbf{h}_{root}$, is sent to the blockchain through the call to the *createNewsStory*(NS_{ID}, h_{root}) smart contract function.
5. The smart contract checks that no other NS has been previously stored in the blockchain with the same $\mathbf{h}_{root}$ and, if so, creates a new entry for this one.
6. Finally, the whole content of T_{NS} is optionally stored in a local or remote database (NSD), for further reference.

The previous process is repeated for every news story, so each news story is contained in a unique tree. Although the news story content will be typically kept in the journal's CMS, our architecture store the full contents in a local database for convenience.

Of course, in a typical news story lifecycle, this content is usually modified, corrected or improved. In these cases, it is necessary to keep an authenticated record of each change to ensure full traceability. Our architecture manages this situation in a very similar way to the certification process, but by calling the *modifyNewsStory()* function of the smart contract instead:

1. Same 1-3 steps than the certification algorithm.
2. Once the construction of $\mathbf{T}_{NS}$ is finished, its identification hash value $\mathbf{h}_{new}$, is sent to the blockchain by calling *modifyNewsStory*(*IDNS, hold, hnew*), where $\mathbf{ID}_{NS}$ is the ID of the new story to be modified, $\mathbf{h}_{old}$ is its previous hash root and $\mathbf{h}_{new}$ the hash of the modified version.
3. Then, this function checks that a new story with ID $\mathbf{NS}_{ID}$ and hash root $\mathbf{h}_{old}$ exists in its internal database, and is not marked as deleted. If so, it adds to the internal state of the new story the modified hash value, $\mathbf{h}_{new}$.

This way, our architecture keeps a full list of the different modifications a news story has suffered over time, knowing the exact time and date the modifications occurred, which provides a complete traceability capability.

This traceability is assured as follows. As can be observed in Fig. 1, the smart contract maintains an internal database for each news story NS. This database is implemented by using the mapping datatype of Solidity, which is essentially a *hash table*, i.e., a collection of tuples (**K**ey, **V**alue). Each entry has the form (NS_i, [h_i^i,...,h_n^i]), where $\mathbf{h}_i^j$ is the hash value corresponding to the *j*-th modification of the news story *i*-th of the system.

This way, the list [h_i^j] contains the history of changes that a specific news story has undergone since its introduction in the architecture. In addition, as each news item is digitally signed, it is also possible to guarantee its authenticity.

C. Verification

Any time later, a third party can verify the integrity of any news story previously published by performing the following steps (see Fig. 3):

1. Reconstruct the Merkle tree $\mathbf{T}_{NS}$ corresponding to the new story **NS,** using the same ordering for the resources that compose **NS**. As a result, a root value $\mathbf{h}_{root}$ is obtained.
2. Call the function *checkStory*(NS_{ID}, h_{root}) of the smart contract, passing $\mathbf{NS}_{ID}$ and $\mathbf{h}_{root}$ as arguments. This function checks if this value exactly matches any of the registers stored in its internal database and, if so, returns a *true* boolean value, along with a timestamp of the certification date of the information. Otherwise, it returns *false.*

As a result, we can obtain the immutability of the information stored in the blockchain (including the smart contract source code). If the result of the previous process is true, the verifier can be sure that the news story was not modified since its publication and certification.

VI. Cost of Using Blockchain for New Stories Tracking

As stated in previous sections, one of the main potential limitations for the integration of blockchain technologies is the cost of data storage

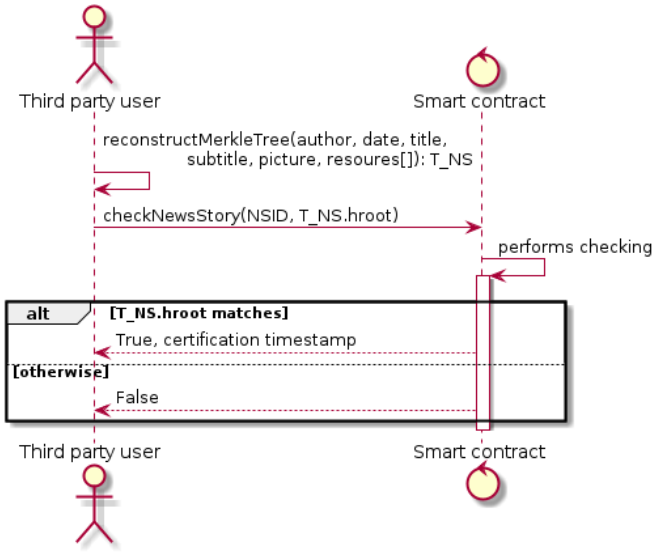


Fig. 3. Sequence diagram to perform the verification process by a third party user of agency.

and transactions processing. Therefore, to estimate the viability of our proposal, it is crucial to properly measure and minimize that cost.

The main reason is related to the execution of smart contracts in blockchains (like Ethereum). In order to reward the nodes that use their computing capacity to maintain the system, each instruction executed requires the payment of a fee in a cryptocurrency (called gas). Simple instructions (such as a sum) cost 1 gas, while others can cost significantly more (e.g., the calculation of a SHA3 hash costs 30 gas). On the other hand, the storage space is especially expensive (around 20k gas for every 256 bits).

There are essentially two approaches to store large volumes of data in public blockchains, which are presented below ordered in terms of complexity (from lower to higher), and economic cost (from higher to lower):

- *Full on-chain storage*: all data is stored, as-is, in the blockchain without any type of pre-processing. For example, news stories could be directly stored as a data structure in a smart contract.
- *Merkle trees*: data is stored off-chain, but it is preprocessed by constructing a Merkle tree structure, which reduces storage costs and increases the bandwidth. Typically, this is combined with data hashing, that is, storing only a hash of the data, in order to guarantee its immutability. The data itself can be stored off-chain in any other system: distributed (e.g., IPFS [15]), cloud, or local (as a common relational database).

In general terms, the storage space in public blockchains is especially expensive compared to computation, in order to discourage its abusive use. For this reason, the use of the first scheme, which is the most inefficient and costly, would commonly imply a prohibitive cost for most uses. As an example, Table II depicts the cost of reading and storing 1 Kilobyte of data in Ethereum in terms of gas units, ether, and US dollars.

TABLE II. Non-volatile Storage Costs In Ethereum. We Have Considered A Gas Price Of 1 Gwei (1 Gwei = 10^{-9} Eth), And 1 Eth = $140 (At The Time Of Writing, January 2020)

Operation	Gas/KB	ETH/KB	$/KB
READ	6,400	0.000032	$0.004
WRITE	640,000	0.0032	$0.448

TABLE III. Cost Calculated Considering 1 Mb Per New Story. Gas And Ether Prices Are The Same As Table II

Scheme	Operation	Cost (Per news story)	Execution time (average)
Both	Smart contract deployment	498274 gas ($0.06972)	19.20 secs
Direct storage	Certification	3.1·108 gas ($448)	10.55 sec
	Modification		
	Deletion	18850 gas ($0.0026)	11.66 sec
	Verification	-	-
Merkle trees	Certification	86848 gas ($0.014)	12.16 sec
	Modification		
	Deletion	18850 gas ($0.0026)	11.66 sec
	Verification	-	-

With these numbers, and considering a typical news story size of around 1MB (including texts and pictures), protecting just a thousand news stories would cost almost half a million dollars. Clearly, this is not a realistic option, discouraged not only in economic terms, but also for security and performance reasons.

For all these reasons, our proposal uses the second approach, based on Merkle trees. This is a much more elegant and efficient solution and, in fact, in most situations, the only viable option. Table III shows all the final results of our experiments.

As can be easily seen, the direct storage scheme is not a viable option in this case, since it would have a prohibitive cost in a real environment. On the other hand, our proposal shows affordable figures, both in economic and performance terms. This is due to the fact that the Merkle tree scheme needs to store only 256 bits per news story, regardless of its total size, which can be arbitrarily large and have any number of internal resources. This way, protecting the integrity of a news story costs around $0.014. This is clearly a very reasonable figure, even for large newspapers with a very large volume of news.

In any case, if necessary, the Merkle tree scheme could be adapted for processing any number of news stories at a fixed cost. The solution implies to create a hierarchy of trees, each containing a unique new story, and by grouping them in any number.

However, this scheme would be considerably more complex to implement and manage than our proposal. In this case, it would be necessary to create and store a (public) cryptographic proof with each new story, that would allow its later verification. This proof should be made available to the verifier.

Regarding the performance, the experiments show that this system is also viable. It is important to note that the tests have been carried out in a testnet, where the confirmation times are higher and have greater variability than in the mainnet. Times have been measured performing each operation ten times, discarding the minimum and maximum times, and calculating the average of the rest.

As can be seen, the execution time is slightly higher than 10 seconds for most of the operations, which seems an acceptable time for the usability of the system. Finally, the retrieval operation, necessary for the verification of a template, is a read-only operation and, therefore, free of cost. In addition, it can be also considered immediate in terms of execution time, due to the fact that the request is processed by the local Ethereum node, and it does not reach the network.

VII. Conclusion and Further Work

To determine a way to address the issues of fake news, disinformation campaigns, and the lack of credibility to which

journalists and media are exposed, throughout in this article we have presented a proposal that uses a blockchain to guarantee the principle of journalistic transparency that enables the tracking and tracing of news stories. This proposal arises as a real implementation alternative that guarantees the fulfilment of this principle, far from being a mere statement of intent by the media.

The proposal detailed in this article still has much room for improvement. For instance, the smart contract could be improved to support several content management systems simultaneously so that news from several agencies or editorial groups related to the same news story can be taken into account.

Another interesting improvement is the use of identities based on elliptic curves so that it is not necessary to use digital signatures dependent on external entities to ensure authorship, although the issue of digital identity from a legal point of view must be solved.

VIII. Annex A. Smart Contract Source Code

```
pragma solidity ^0.5.11;
pragma experimental ABIEncoderV2;

contract NewsAuth {

  struct NewsStory{
    uint256[] entries;
    bool exists;
  }

  /// Each template is indexed by an user ID
  mapping(uint => NewsStory) NS_database;


  /// Store a new template
  function createNewsStory(uint NS_ID,
                           uint256 h_root) public {
    // Check that there is no news story with this ID
    require(NS_database[NS_ID].exists != true,
            "News story already exists.");

    uint256[] memory _entries = new uint256[](1);
    _entries[0] = h_root;

    /// Add news story to internal database
    NS_database[NS_ID] = NewsStory({
          entries: _entries,
          exists: true
    });
  }

  /// Modify a user template
  function modifyNewsStory(uint NS_ID,
                   uint256 h_new) public returns (uint) {
    // Check that there is no news story with this ID
    require(NS_database[NS_ID].exists == true,
            "News story does not exist.");

    // Add modification to NS's history
    return NS_database[NS_ID].entries.push(h_new) -1;
  }

  /// Return an specific template
  function deleteNewsStory(uint NS_ID) public {
    delete NS_database[NS_ID];
  }

  /// Return a user template
  function checkNewsStory(uint NS_ID,
           uint256 h_check) view public returns (bool) {
    // Check that there is no news story with this ID
    require(NS_database[NS_ID].exists == true,
            "News story does not exist.");

    bool flag = false;
    for (uint i = 0;
         i < NS_database[NS_ID].entries.length;
         i++) {
      flag = (NS_database[NS_ID].entries[i]==h_check);
      if(flag == true)
        break;
      }
      return flag;
  }
}
```

Acknowledgment

This research work has been funded by the Madrid Regional Government through the project e-Madrid-CM (P2018/TCS-4307). The e-Madrid-CM project is also co-financed by the Structural Funds (FSE and FEDER).

References

[1] B. McNair, *Fake News: falsehood, fabrication and fantasy in journalism*. London: Routledge, 2017.

[2] D. M. J. Lazer et al., "The science of fake news," *Science*. vol. 359, no. 6380, pp. 1094-1096, 2018.

[3] A. Phillips, "Transparency and the new ethics of journalism," *Journal. Pract.*, vol. 4, no. 3, pp. 373–382, Aug. 2010.

[4] M. Karlsson, "The immediacy of online news, the visibility of journalistic processes and a restructuring of journalistic authority," *Journal. Theory, Pract. Crit.*, vol. 12, no. 3, pp. 279–295, Apr. 2011.

[5] M. Revers, "The twitterization of news making: transparency and journalistic professionalism," *J. Commun.*, vol. 64, no. 5, pp. 806–826, Oct. 2014.

[6] K. Chadha and M. Koliska, "Newsrooms and transparency in the digital age," *Journal. Pract.*, vol. 9, no. 2, pp. 215–229, Mar. 2015.

[7] T. P. Vos and S. Craft, "The discursive construction of journalistic transparency," *Journal. Stud.*, vol. 18, no. 12, pp. 1505–1522, Dec. 2017.

[8] M. Karlsson and C. Clerwall, "Transparency to the rescue?," *Journal. Stud.*, vol. 19, no. 13, pp. 1923–1933, Oct. 2018.

[9] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[10] D. Lasorsa, "Transparency and other journalistic norms on twitter," *Journal. Stud.*, vol. 13, no. 3, pp. 402–417, Jun. 2012.

[11] L. Morton, "Where are you coming from?," *Journal. Pract.*, vol. 9, no. 2, pp. 168–183, Mar. 2015.

[12] S. O'Riordan, G. Kiely, B. Emerson, and J. Feller, "Do you have a source for that?," in *Proceedings of the 15th International Symposium on Open Collaboration*, 2019, pp. 1–10.

[13] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system." 2009.

[14] A. L. Franzoni, C. Cardenas, and A. Almazan, "Using Blockchain to store teachers' certification in basic education in Mexico," in 2019 *IEEE 19th International Conference on Advanced Learning Technologies* (ICALT), 2019, pp. 217–218.

[15] J. C. Cheng, N. Y. Lee, C. Chi, and Y. H. Chen, "Blockchain and smart contract for digital certificate," in *Proceedings of 4th IEEE International Conference on Applied System Innovation* 2018, ICASI 2018, 2018.

[16] B. Notheisen, J. B. Cholewa, and A. P. Shanmugam, "Trading real-world assets on blockchain," *Bus. Inf. Syst. Eng.*, vol. 59, no. 6, pp. 425–440, Dec. 2017.

[17] Y. Yuan and F.-Y. Wang, "Towards blockchain-based intelligent transportation systems," in 2016 *IEEE 19th International Conference on Intelligent Transportation Systems* (ITSC), 2016, pp. 2663–2668.

[18] R. Khalil, A. Zamyatin, G. Felley, A. Gervais, and P. Moreno-sanchez, "Commit-chains : secure, scalable off-chain payments," *Cryptol. ePrint Arch.*, p. 642, 2018.

[19] K. M. Khan, J. Arshad, and M. M. Khan, "Secure digital voting system based on blockchain technology," *Int. J. Electron. Gov. Res.*, vol. 14, no. 1, pp. 53–62, Jan. 2018.

[20] R. Hanifatunnisa and B. Rahardjo, "Blockchain based e-voting recording system design," in 2017 *11th International Conference on Telecommunication Systems Services and Applications* (TSSA), 2017, pp. 1–6.

[21] D. Tse, B. Zhang, Y. Yang, C. Cheng, and H. Mu, "Blockchain application in food supply information security," in 2017 *IEEE International Conference on Industrial Engineering and Engineering Management* (IEEM), 2017, pp. 1357–1361.

[22] H. M. Kim and M. Laskowski, "Towards an Ontology-Driven Blockchain Design for Supply Chain Provenance," *SSRN Electron. J.*, 2016.

[23] H. Kaur, M. A. Alam, R. Jameel, A. K. Mourya, and V. Chang, "A proposed solution and future direction for blockchain-based heterogeneous medicare data in cloud environment," *J. Med. Syst.*, vol. 42, no. 8, p. 156, Aug. 2018.

[24] R. C. Merkle, "A digital signature based on a conventional encryption function," in Conf. on the Theory and Applications of Cryptographic Techniques on Advances in Cryptology (CRYPTO), 1988, pp. 369–378.

[25] K. Mouratidis, D. Sacharidis, and H. Pang, "Partially materialized digest scheme: an efficient verification method for outsourced databases," *VLDB J.*, vol. 18, no. 1, pp. 363–381, Jan. 2009.

[26] H. B. Ribeiro and E. Anceaume, "DataCube: A P2P persistent data storage architecture based on hybrid redundancy schema," in 2010 *18th Euromicro Conference on Parallel, Distributed and Network-based Processing*, 2010, pp. 302–306.

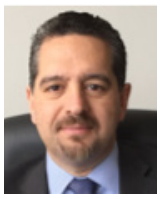
[27] C. Dannen, *Introducing Ethereum and Solidity: Foundations of cryptocurrency and blockchain programming for beginners. Berkeley*, CA, USA: Apress, 2017.

Francisdo Jurado

Francisco Jurado is Lecturer in the Computer Engineering Department of the Universidad Autónoma de Madrid, Spain. He received his Ph.D. degree with honors in Computer Science from the University of Castilla-La Mancha in 2010. His research areas include Natural Language Processing, and Computer Supported Collaborative Environments.

Oscar Delgado

Oscar Delgado received the B.S. degree in Computer Science from the Universidad Politecnica in 2002, and the PhD in Telecommunications Engineering by Universidad Carlos III de Madrid in 2011. His research interests include cryptology, network security, cryptocurrencies and blockchain. He leads the chair on blockchain technologies at Universidad Autonoma de Madrid funded by Grant Thornton.



Alvaro Ortigosa

Alvaro Ortigosa was born in San Carlos de Bariloche, Argentina, in 1968. He holds a Ph.D. on Computer Science from the Universidad Autónoma de Madrid in 2000, a M.S. on Computer Science from the Universidade Federal de Rio Grande do Sul in 1995 and a degree on Computer Science from the Universidad Nacional del Centro de la Prov. de Buenos Aires in 1993. He is director of the Research Institute for Forensics and Security Science of UAM since 2017 and Associate Professor at the Department of Computer Science of UAM since 2001. His main research lines are adaptive systems and user modeling, application of datamining for user model acquisition, personality and emotion detection through text and virtual social network analysis, and application of datamining to risk analysis. He has (co)authored more than 60 papers in international journals and conferences. Mr. Ortigosa is member of European Cybercrime Training and Education Group.

Traceable Ecosystem and Strategic Framework for the Creation of an Integrated Pest Management System for Intensive Farming

Miguel Ángel López, Juan Manuel Lombardo*, Mabel López, David Álvarez, Susana Velasco, Sara Terrón

Fundación I+D del Software Libre (FIDESOL), Granada (Spain)

Received 27 April 2020 | Accepted 29 July 2020 | Published 7 August 2020

Abstract

The appearance of pests is one of the major problems that exist in the growth of crops, as they can damage the production if the appropriate measures are not taken. Within the framework of the Integrated Pest Management strategy (IPM), early detection of pests is an essential step in order to provide the most appropriate treatment and avoid losses. This paper proposes the architecture of a system intensive farming in greenhouses featuring the ability to detect environmental variations that may favour the appearance of pests. This system can suggest a plan or treatment that will help mitigate the effects that the identified pest would produce otherwise. Furthermore, the system will learn from the actions carried out by the humans throughout the different stages of crop growing and will add it as knowledge for the prediction of future actions. The data collected from sensors, through computer vision, or the experiences provided by the experts, along with the historical data related to the crop, will allow for the development of a model that contrasts the predictions of the actions that could be implemented with those already performed by technicians. Within the technological ecosystems in which the Integrated Pest Management systems develop their action, traceability models must be incorporated. This will guarantee that the data used for the exploitation of the information and, therefore for the parameterization of the predictive models, are adequate. Thus, the integration of blockchain technologies is considered key to provide them with security and confidence.

Keywords

Integrated Pest Management, IPM, Computer Vision, Machine Learning, Artificial Intelligence, Blockchain, Intensive Farming, Greenhouses.

DOI: 10.9781/ijimai.2020.08.004

I. Introduction

PRESERVING the crops and increasing production is a constant concern of vegetable growers. The appearance of diseases and external agents such as fungi, bacteria, insects, and other forms of plant or animal life can be a threat that jeopardizes the crops, and hence the livelihood of growers and their families. This does not only put the primary sector groups at risk, but it also has a direct impact on the economy of a country and compromises the nutritional security of the inhabitants of the countries that are dependent on the supply of such goods.

In order to counter the various pests that endanger the supply of food and agricultural commodities, a wide range of products called agrochemicals, phytosanitary products or pesticides are used. These highly toxic synthetic compounds include chemical components that can eliminate pests and vectors of communicable diseases while avoiding damage to the crops [1], [2].

Given the controversy over the permissiveness in the use of certain chemicals, which are banned in some countries [3] and exploited in others [4], epidemiological studies show that human exposure to these agrochemicals may produce adverse effects on health [5]–[10].

In addition, the excessive or indiscriminate employment of agrochemicals can lead to a decrease in their own efficiency, generating just the opposite effect to the one desired [11], [12]. These substances can accelerate the development of resistant weeds and favour the increase of the pest population. Such resistance is more likely to occur in genetically modified crops, as they tolerate a very narrow spectrum of herbicides [13]. This makes rotation difficult due to the depletion of the biodiversity of agricultural land [14], which causes the development of persistent weeds, insect pests and pathogenic bacteria that evolve and spread rapidly [15]. The latter may lead to use more toxic profile herbicides [16] that will most probably turn detrimental to the environment [17]. The grower gets trapped in a harmful and unsustainable circle given that those varieties require a greater contribution of fertilisers, water and pesticides, which increases the dependence on the use of synthetic products and may lead to a decrease in production [18], [19].

Adding to the drawbacks implied in the application of the products mentioned above, such agricultural waste is the most widespread, dispersed and difficult to control, and becomes one of the main sources of contamination of groundwater, surface water and soil. This serious environmental problem is also transferred to the population of developed countries where the problems of food poisoning have also had a negative social and economic impact [20], [21].

* Corresponding author.
E-mail address: jmlombardo@fidesol.org

The desire of consumers to obtain food free of harmful chemicals, together with the need for a real system of sustainable and organic farming, has triggered the creation of an Integrated Pest Management system (IPM), as a scientific paradigm in safe food production. This procedure allows to manage the agroecosystem favouring the grower through the observation of environmental aspects that enables the incorporation of alternative non-chemical measures or techniques to pesticides. Integrated Pest Management "emphasises the growth of a healthy crop with the least possible disruption to agro-ecosystems and encourages natural pest control mechanisms" [22].

The development of populations of harmful organisms is avoided by the preventive application of those control measures, which minimizes risks to both human health and the environment. Therefore, it is essential to accurately detect different pest infestations when they are just emerging and to establish a control and prevention strategy that is as effective as possible. In this sense, it is essential to identify the problem correctly. However, carrying out these actions involves several difficulties. An adequate diagnosis of the situation relies on historical data obtained about the area, the knowledge of growers, or the observations and sampling conducted by technicians on the crop areas [23] on a weekly or biweekly basis [24].

Furthermore, global trade and climate change favour the appearance of new species and the displacement of habitats, making the future spread of harmful organisms unpredictable [25]. Accordingly, early detection of pests and the minimum application of plant protection products are critical requirements to avoid an increase in pollution and obtain less chemically treated though healthier agricultural products.

In response to these demands, different technological solutions have emerged for the detection of pests using intelligent aircraft. Some studies, such as that of Hunt et al. [26], propose a pest detection system based on sensors and unmanned aircraft systems to acquire high-resolution images of potato crops. To check if a crop is infected, these authors analyse the nitrogen content in the leaves of the plants, allowing them to act quickly. In this regard, Stumph et al. [27] evaluate an automatic system for the early detection of pests that uses unmanned aerial vehicles equipped with cameras too, ultraviolet technology and algorithms based on computer vision. Likewise, Shamshiri et al. [28] study the behaviour of the pests in oil palm plantations drawing on a drone equipped with sensors. Its aim is the design of three controllers (PID, LQR, and LQR plus observer) that allow for automatic insects detection. Finally, Psirofonia et al. [29] carry out a study on three cases where a drone, in combination with electronic traps, is used for pest detection and spraying of olive and palm trees.

In other studies, researchers also use computer vision techniques and a sensor system to detect pests in greenhouse crops [30]–[32]. Huddar et al. [33] propose an algorithm for pest detection based on image processing that has be tested on outdoor crops as well as in greenhouses. Their algorithm is 96% accurate in detecting whiteflies, one of the most damaging pests to crops. Studies such as that of Nguyen and Nansen [34] use data provided by hyperspectral sensors to feed an algorithm that classifies the risk of pest infestation in spinach and cabbage plantations.

In this work, we present the architecture of an automatic system based on sensors and cameras for the identification, recording, and tracking of different pests that may affect intensive farming crops in greenhouses. In section II, the proposed architecture and its components are defined and explained in detail. The discussion about this architecture is exposed in section III. Finally, the conclusions and future work are considered in section IV.

II. Material and Methods

Our research aimed at creating a system that is able to autonomously monitor the state of crops, analyse their evolution, enable early detection of pests, and forewarn the emergence of new ones as well as their likely transmission vectors.

The design shown corresponds to an Integrated Pest Management system harnessing the capacity to collect information from various sources, analyse such data, extract characteristics, or draw patterns for a better understanding of the behaviour of external agents that may affect the development of crops. Among these external agents are mainly insects or pests, and all those climatic, environmental or biological factors influencing the growth of plants. In this way, a predictive plan can be created to enable preventive action, which will reduce the application of agrochemicals significantly and will contribute to the long-term sustainability of the agri-food system.

This section describes the IPM strategies and techniques that have been applied to determine each of the system components for the intelligent acquisition, processing, and evaluation of the information required to model the system architecture. The data provided and their sources, which are also essential to create our predictive pest management model, are considered.

Research conducted by Fidesol has determined that the parameters described in Table I are the most appropriate ones to be considered when addressing an IPM system for intensive crops farming.

TABLE I. Influential Parameters in the Appearance of Pests and Other Elements Harmful to Plants

Parameter	Description
Soil temperature (ºC)	Soil temperature at root level
Soil moisture (%RH)	Soil moisture at root level
Soil electrical conductivity (s/m)	Soil electrical conductivity at the root level
Solar radiation (W/m2)	Direct and diffuse radiation
Air temperature (ºC)	Air temperature
Air humidity (%RH)	Relative air humidity
Precipitation (mm/h)	Water content in precipitation
Altitude (m)	The height of the area in relation to sea level
Atmospheric pressure (Pa)	Atmospheric pressure outside the greenhouse
Wind direction (º)	Wind direction
Wind speed (m/s)	Wind speed
Surface leaf temperature (ºC)	Surface leaf temperature
Stem thickness (mm)	Stem thickness
Nitrates (ppm)	Nitrate levels in the soil
Potassium (ppm)	Potassium level in the soil
CO2 (ppm)	CO2 level in the air
Soil tension (kPa)	The effort that the roots have to make to extract the water they need from the soil
PH (pH)	PH level of the soil
Soil type	Soil type classification
Property	Property where the cultivation takes place
Agricultural area	Greenhouse where the cultivation takes place
Plant genus	Type of plant
Seed variety	Varieties of seeds
Grafting	Kind of graft
Plantation type	Transplant form
Row distance (m)	Distance between rows
Plants distance (m)	Distance between plants
Cycle	Time cycle when cultivation takes place
Previous sowing	Plant previously sown
Sample date	Date on which the sample is collected

A. IPM Strategies and Techniques

The analysis of the state of the plants, together with the study of the growth rate of the pest populations is intended to develop control strategies aimed at reducing the use of chemical controls, through activities of surveillance, biological and chemical control.

The main application of IPM took place in agriculture, although the benefits of its implementation in other fields have been proven later, such as in the control of weeds, diseases, and other plagues that can damage crops, gardens and other urban spaces [35]–[37].

With the adoption of the IPM, activities are now focused on preventive monitoring and analysis of crops to reduce the populations of harmful organisms, using appropriate cultivation techniques and ecological measures. The IPM prioritizes non-chemical methods so that professional pesticide users opt for practices and products that pose the least risk to human health and the environment. Therefore, after a regular inspection on the appearance of a specific pest, its growth is observed and, if it exceeds a certain threshold, the action is executed by applying a preferably biological treatment [22]. Only when the control results are not satisfactory, the appropriate plant protection products are applied for each case. A general outline of the IPM system is shown below in Fig. 1.

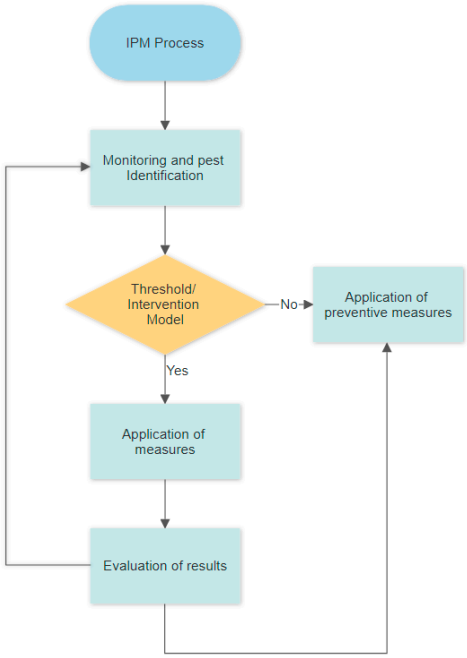


Fig. 1. General outline of the steps identified for the IPM system.

In this research work, for the adoption and correct compliance of IPM strategy, the following steps are considered to achieve the adequate environmental sanitary conditions and avoid the proliferation of harmful organisms:

- Monitoring and identification of pests. It is worth mentioning that it is not necessary to reduce the population of all insects that may appear in pest control. Some of them are harmless and even beneficial. It is essential to identify and differentiate them in order to act according to the threshold defined within the framework for action agreed at the national level. Through monitoring and identification, the possibility of unnecessary or incorrect use of pesticides is removed.
- Checking the action threshold / intervention model. Before taking any control action, it should be observed whether a threshold has been reached where conditions indicate that pest control measures should be taken.
- Action/application of control measures. Action is taken when the pest has been monitored and identified, and the control thresholds have been exceeded even following the application of preventive methods that have not proved effective enough in controlling the pests and diseases. The application of cultural, physical or biological methods, which pose the least risk to health and the environment, such as trapping or weeding, always comes first in the implementation of control measures. If, after application of the relevant measures, the results continue to exceed the action thresholds, a switch to the use of chemicals shall be made by selecting those authorised for effective control. These chemicals should be as much compatible as possible with non-target organisms to avoid harm to insects for natural pest control and others that are also beneficial, such as bees [38].
- Evaluation of results. At this step, the results of the prevention or control actions taken are assessed in order to know their degree of effectiveness or unsought effects, which will help to determine the appropriate actions if a similar case arises in the future. In addition, a report will be produced whose main purpose is to verify the degree of effectiveness of the measures, identify trends through the analysis of results, and include the observations deemed necessary.
- Prevention. One of the first actions of pest control in IPM strategy is the application of preventive measures to prevent pests from becoming a threat. Some of these measures are based on the following methods: crop rotation, the use of appropriate cultivation techniques, or the selection of crop species resistant to pests. These methods are usually profitable in the long-term as well as friendly with humans and the environment [39]. This type of measures builds the framework of the first actions to be implemented for the prevention or elimination of harmful organisms.

Each of the steps that make up the IPM have been considered in the architecture design proposed in this work, as detailed in the following section.

B. Proposed Architecture

The architecture proposed in this work is defined within the framework of Directive 2009/128/ EC on Sustainable Use of Pesticides of the 21 October 2009, based on the principles of the IPM that consider "all available plant protection methods and subsequent integration of appropriate measures that discourage the development of populations of harmful organisms and keep the use of plant protection products and other forms of intervention to levels that are economically and ecologically justified and reduce or minimize risks to human health and the environment" [22].

Once the correspondence of the approaches and measures of the IPM strategy with the necessary technological components that promote the application of its principles has been identified, the system architecture is described in Fig. 2.

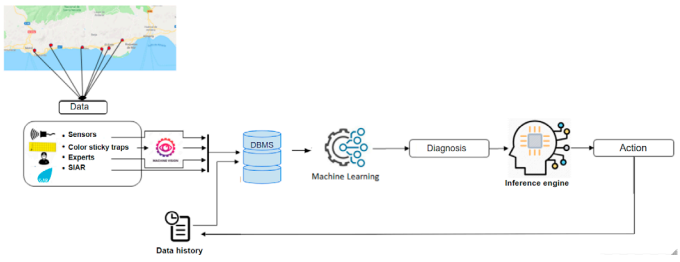


Fig. 2. System architecture.

By relying on the data provided continuously by the different sources and also on the data history, the proposed system is able to early detect incidents that may occur during the life of a crop and

determine the physiological state from its phenological stage in real-time. All this information flow helps to establish a sound foundation on which to create a predictive system based on machine learning.

However, the implementation of an Integrated Pest Management system is particularly challenging since the information for each plant is not yet fully defined, as is the case with thresholds for all types of pests. There is a need for plant health expertise to help ensure the sustainable and efficient use of plant protection products. The determination of safe and scientifically sound threshold values depends on the place of cultivation and is even a controversial issue among experts [40]. The impressive amount and complexity of variables that intervene in this system to define parasitic phenomena (both of environmental origin and of the crop itself) make it necessary to carry out well-oriented and advanced research in which technology and fieldwork are highly relevant. In addition to the parameters considered in the system described here and the knowledge acquired during the learning process of the system itself for this environment over several complete production cycles, the presence of an infestation will be essential to detect accurately its appearance in advance, as well as its level of severity so that the most suitable solution can be adopted in each case.

Regarding the steps that make up the IPM system, the guidelines to create the proposed architecture are explained below.

1. Monitoring and Identification of Pest

To enhance the prevention principle of Integrated Pest Management, monitoring and identifying the species that can negatively affect the production of a given crop is essential. In this case, monitoring typically entails a visual process that implies exploring and observing the plants on a regular basis. This type of observation is a time-consuming activity that may result in the pest not being timely detected, but just when the infestation has reached very high levels. The aim of this research work is the early detection of pests through the creation of an intelligent system featuring the ability to monitor and identify the level of infestation by resorting to data acquisition techniques specific to precision agriculture and computer vision.

a) Data Acquisition

In order to know the status of a crop, it is essential to collect data for subsequent analysis and act on the results obtained. Automatic data collection is a critical step as this generate a more specific view of all the elements that can affect a crop.

- **Technical work.** The monitoring process is usually performed manually by qualified technicians. This task requires checking the plates used to detect crop pests for at least two hours a day. So, this is not always possible if other more critical tasks arise. For the horticultural plants in the study area, the placement of the plates should be one every 100 m2, i.e., 100 plates in 1 hectare. The inspection is conducted once or twice a week to examine the species caught and check the growth or decline of the population observed while taking into account the threshold for action. Should this activity be reinforced with automation techniques, better monitoring, early detection and more frequent assessment of the status of the crop will be available.
- **Wireless sensor network**. It allows to obtain continuously the parameters associated with the crop in real-time, providing greater control of the exploitation and capacity in case of contingencies. It consists of a remote unit with an autonomous power supply that includes a solar panel and battery together with sensors that collect and transfer the environmental information (climate and soil parameters) to other remote unit. The continuous collection of such information becomes a key component of a more accurate early detection mechanism to respond quickly to adverse situations.

  It is paramount to take into consideration the implementation of information assurance technologies, both on input devices and sensors. Internet of Things (IoT) devices pose a security challenge since well-established data protection protocols are not yet comprehensively embedded and there is no a "one-size-fits-all" solution. For IoT to be fully exploited, information and communication encryption along with appropriate authentication requests must be implemented. Blockchain technology allows to store encrypted information for each IoT device. A complete security layer will also be provided in the blockchain IoT system so that authentication will be more secure. Blockchain will allow for security and non-repudiation management in IoT sensor networks and the certification of multimedia content assuring the veracity, transparency and security of the information.
- **Data mining of external systems**. Agricultural information systems are of particular interest as they make it possible to identify a large number of properties, their characteristics, data history, yields, pests, common diseases, monitoring performed, and control measures taken. This type of instrument enables agricultural technicians to record information and observations on the plot. The work of the technicians is time demanding, albeit it is essential since the historical data on the crop resulting from monitoring activities are invaluable for decision-making on the integrated management pest strategy.
- **Data from the Agroclimatic Information System (SIAR)** [41]. The SIAR network "has more than 460 stations located on irrigated areas in 12 autonomous communities in Spain. This collaborative network has a coverage of 90% of the irrigable area of the country and allows obtaining meteorological data with an hourly average frequency from one or several stations" [42]. The parameters obtained through this source of information provide value not only in terms of temperature and humidity but also for wind direction and intensity. As a consequence, it is possible to study the spread of a pest on a small (in the same agricultural area) and a large geographical scale (different agricultural areas).
- **Computer Vision Systems**. Within an automatic identification system, computer vision plays a fundamental role since it allows for the recognition of objects just as the human eye could. The use of this methodology has great potential in identifying insects based on size, shape or colour. In addition, this system is more effective if it describes at what development stage the insects tested are in order to know the status of the pest and act accordingly.

b) Register

Data sources will be dealt with NoSQL and SQL storage systems for the management of large volumes of unstructured and structured data according to their purpose, so as to reduce query latency in real-time information requests.

c) Sensor Data Normalization

Data normalization is an essential step for the integration of devices from different manufacturers and therefore offer different data structures. The normalization of the data sent by each device aims to obtain a standard scale in the values, reduce the data structures and achieve standardization, thus avoiding the loss of information and processing errors (Fig. 3). The aim is to transform data into a standard format that allows the exchange and representation of information from different heterogeneous devices [43]. This technique allows us to prepare the data so that the algorithms can model them properly. A difference in the scale of values can cause problems when combining data as characteristics in a model. By fitting in the same distribution within a series of calculations, a more stable behaviour in optimization methods is achieved.

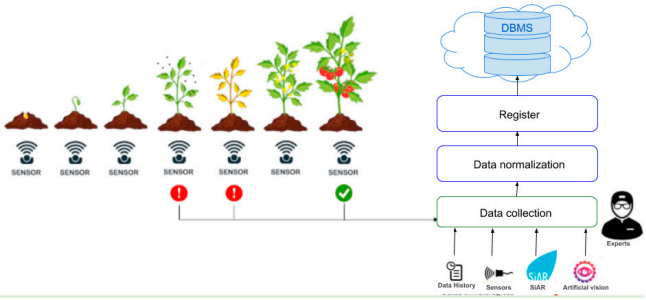


Fig. 3. Monitoring and identification of pests by the system.

d) Processing

The analysis of the characteristics of the data sets is carried out to limit them and thus avoid working with repeated and irrelevant data. To that purpose, structures are simplified in terms of their dimension and complexity, which improves understanding of the information and performance. The reduction of this dimensionality is also aimed at the segmentation and classification of the data with supervised classification algorithms. This action focuses on avoiding processing errors in the algorithms for which the review is performed in order to eliminate erroneous data by completing the information in the event of sensor failures and generating data that can be verified from nearby stations.

2. Diagnosis According to Thresholds and Application of Prevention and Control Measures

Once the information from the different sources has been collected, the anomaly monitoring and identification module can provide an intelligent diagnosis that will be reviewed in the integrated pest management plan and action will be taken, if necessary. The results of such monitoring shall be based on the threshold levels established for each harmful organism in the given region, the specific areas, the crops, the climatic conditions of each plot, and the level of economic damage.

The sighting of a pest does not always imply the need for control since it will not be necessary to take a limiting action if it does not exceed the established threshold of infestation. An inference engine is responsible for making intelligent decisions basing on the knowledge already acquired or on new data registered so that the system feeds back to infer an appropriate diagnosis and determine an action according to the circumstances analysed. The methodology used by the inference engine follows the process presented in Fig. 4.

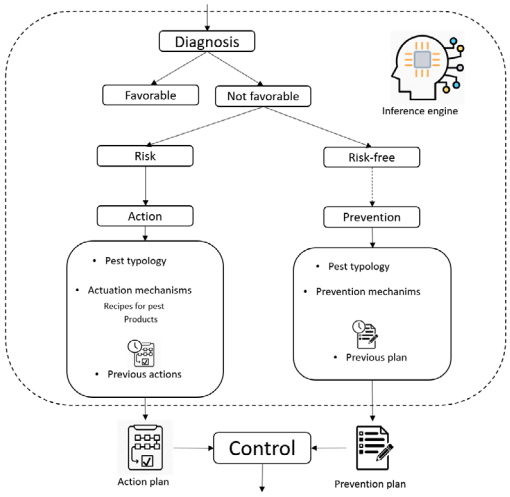


Fig. 4. Inference engine. Evaluation of the input for diagnosis and action or prevention plan.

3. Evaluation

Professional users shall check the effectiveness of the measures implemented on the basis of recorded data on prevention, control and monitoring of harmful organisms. All the protection methods applied and the results obtained shall be carefully examined for subsequent introduction into the system (either manually or automatically). The threshold values of the parameters above to which the different prevention and action measures are applied will be reviewed to adjust them as much as possible when the problem that caused them has already been solved.

In addition, the information in the system will be updated with the registration of all the variations in terms of products authorised and expressly approved for cultivation, new developments in innovative products that are healthy and environmentally friendly, as well as with the alterations or additions to legal and health matters. This also ensures that the product meets the quality requirements set.

The large amount of information and learning are processed in a cognitive engine that feeds the inference engine so that the latter can make intelligent decisions based on the knowledge already acquired or on new data. So, the system feeds back to infer a diagnosis that is increasingly accurate and adapted to the circumstances analysed.

III. Discussion

IPM is a fundamental part of all activities involved in the efficient management of a greenhouse. In this system, the detection and classification of insects is a must. The employment of various automatic methods for pest detection in the crop optimizes this task and significantly improves the results.

One of the data collection methods in the architecture proposed in this work begins with insect sampling to assess the status/level of infestation. Some of the research in this area shows the potential of image processing algorithms to count insects captured in traps. In one of the works analysed, a system based on colour heuristics and mathematical morphology is used [44]. This system consists of the delimitation of the region of interest, application of the colour transformations, segmentation based on thresholds, and detection of young whitefly nymphs. Others rely on simple image processing systems, using size, colour, or outline to detect specimens [45]–[47]. It was shown that the count was done faster automatically than manually. Some authors propose a method for identifying whiteflies in greenhouses based on their own algorithm called LOSS [48]. However, this method is later improved by abstracting the position and illumination to achieve a better identification process [49]. Also noteworthy are the works for the detection of aphids [46], [50] and whiteflies [51] in greenhouses using different formulas with which a system based on Support Vector Machines (SVM) will make the classification.

In our project, with the module of pest recognition by computer vision, we try to detect the spectrum that comprises the main harmful organisms causing higher economic losses in agricultural production. The diagnosis of pests will be obtained by studying the level of occupation in the leaves of the plants and their symptoms, by means of the most efficient algorithms of classification based on SVM, as they are the ones providing the best results, which will be evaluated to obtain the highest possible efficiency for each of the insects.

However, it should be noted that different computer vision techniques may not be sufficient for early and high-quality detection. The project incorporates self-adjusting algorithms for acquiring knowledge using machine learning techniques, viewed from a perspective based on agricultural knowledge and the capacity for critical evaluation acquired from the observation of crops development.

In this line and with a multidisciplinary approach that involves

the observation of plants, the interpretation of images obtained from cameras, and decision-making for early pest detection, research such as that of Boissard [52] focuses on the application of image processing algorithms adapted with machine learning techniques. Some studies use SVM-based algorithms as a classification method [51], [53], [54], cascaded convolutional neural networks [55] or neural learning algorithms together with knowledge-based techniques for an automatic image interpretation system [56]. The system proposed goes further since the learning algorithms used to obtain a pest diagnosis not only act in the computer vision system but also consider the whole set of standardized input data that have been collected from different sources of information in the system.

Recent works use additional components from the precision agriculture that integrate a network of sensors to measure temperature, humidity and light intensity levels in a greenhouse, using image acquisition and computer vision system [57]. Also, in order to control pests, a system of light, relative humidity and soil moisture sensors are usually installed [58]–[61], or underground acoustic sensor nodes with a mathematical simulation model [62].

In the work of Calderón [61], the variables related to environmental conditions such as air temperature, relative humidity, CO2 concentration, luminosity, wind direction and speed, solar and ultraviolet radiation are studied for monitoring. Likewise, the analysis of soil moisture in multiparametric monitoring for this product is considered of interest [63]. On the other hand, regarding the study of fungal diseases, air temperature and humidity are observed as critical factors [64]. All these parameters have been considered when designing the architecture proposed in this work, together with others also considered such as edaphological, meteorological and indoors environmental parameters, and those relating to the state and characteristics of the plant variety, as well as the location in the greenhouse. As a result, we have more than thirty parameters that allow for a very precise pest detection and control framework that can even help to find new variables that influence the appearance of pests and diseases in crops.

IV. Conclusions and Future Work

Agriculture is at a turning point as it is facing new critical challenges. There is no doubt about the growing need to increase production in order to meet the demands for food. There is also a need to improve production methods and techniques, and obtain higher quality products. In this sense, society demands the integration of agricultural practices that respect human health and the environment, and favour the preservation of heritage for future generations while guaranteeing food security [65].

Integrated Pest Management enables the development of healthy crops by maintaining agroecosystems and promoting pest control mechanisms. The relevance of this system stems from the benefits it provides and its mandatory nature for professional users as far as plant protection products are concerned.

In this paper, our research project has been presented as a system that follows and applies the principles of the IPM. The steps for the implementation of this system are described, and its principles are analysed through the integration of the necessary hardware and software components in each of the steps of the IPM. This allows to build an intelligent system for the early prediction and quick response on intensive farming in greenhouses.

The main contribution of this work is the development of an automated, secure and integrated system that can provide accurate information on pests affecting plants, predict their appearance and facilitate, through a system of recommendations, the necessary preventive and control measures to be implemented in an effective and environmentally friendly way. By considering environmental factors, appearance and insect activity, it is possible to predict the appearance of pests and diseases in an uncontrolled agricultural environment and apply the most appropriate measure in each case.

Work will continue in the search for the most effective algorithms that allow for the diagnosis of diseases and pests through the large volume of parameters provided on a continuous basis along with the building of the knowledge base that will feed the system. It will be necessary to identify the limitations imposed by the presence of elements that disrupt and alter the measurements of the system, e.g. the presence of dirt, excessive light or silence in the data acquisition system that have provided erroneous data in the diagnosis provided. Finally, the study of the behaviour of pests and diseases will be extended so that the proposed system will become more accurate across different experimental locations resulting from the comparison and improved results from real situations. In this way, new influencing factors can be identified, and those that are more unpredictable and represent a threat to the development of both this and any other type of crop can be pointed out.

References

[1] O. Díaz and C. R. Betancourt Aguilar, “Los pesticidas; clasificación, necesidad de un manejo integrado y alternati- vas para reducir su consumo indebido: una revisión,” Revista Científica Agroecosistemas, vol. 6, no. 2, pp. 14–30, 2018.

[2] A. Ferrer, “Intoxicación por plaguicidas,” Anales del Sistema Sanitario de Navarra, vol. 26, no. SUPPL. 1, pp. 155–171, 2003.

[3] European Comission, Regulation (EU) No 540/2011 of 25 May 2011 implementing Regulation (EC) No 1107/2009 of the European Parliament and of the Council as regards the list of approved active substances. 2011.

[4] K. Zhang et al., “Susceptibility of Sogatella furcifera and Laodelphax striatellus (Hemiptera: Delphacidae) to Six Insecticides in China,” Journal of Economic Entomology, vol. 107, no. 5, pp. 1916–1922, 2014.

[5] A. Samsel and S. Seneff, “Glyphosate, pathways to modern diseases IV: Cancer and related pathologies,” Journal of Biological Physics and Chemistry, vol. 15, no. 3, pp. 121–159, 2015.

[6] M. F. Fernández, M. F. Fernández, J. A. López-Medina, V. Mustieles, and N. Olea, “Obesógenos ¿Una nueva amenaza para la salud pública?,” Revista de Salud Ambiental, vol. 17, no. 1, pp. 93–99, 2017.

[7] R. Gerber et al., “Bioaccumulation and human health risk assessment of DDT and other organochlorine pesticides in an apex aquatic predator from a premier conservation area,” Science of the Total Environment, vol. 550, pp. 522–533, 2016.

[8] A. Sabarwal, K. Kumar, and R. P. Singh, “Hazardous effects of chemical pesticides on human health–Cancer and other associated disorders,” Environmental Toxicology and Pharmacology, vol. 63, pp. 103–114, 2018.

[9] E. Nava-Pérez, C. García-Gutiérrez, J. R. Camacho-Báez, and E. L. Vázquez-Montoya, “Bioplaguicidas: Una opción para el control biológico de plagas,” Ra Ximhai, vol. 8, no. 3, pp. 17–29, 2012.

[10] A. Samsel and S. Seneff, “Glyphosate, pathways to modern diseases III: Manganese, neurological diseases, and associated pathologies,” Surgical Neurology International, vol. 6, no. 1, 2015.

[11] A. B. Figueras, “Representación social del riesgo en la costa de Oaxaca: Agrotóxicos, salud y medio ambiente,” Arxiu d’Etnografia de Catalunya, pp. 115–149, 2020.

[12] S. Ramírez, “Las investigaciones de salud pública en Latinoamérica. Reflexiones desde el Sur global,” Revista Facultad Nacional de Salud Pública, vol. 37, no. 1, pp. 106–113, 2019.

[13] S. Bonny, “Genetically Modified Herbicide-Tolerant Crops, Weeds, and Herbicides: Overview and Impact,” Environmental Management, vol. 57, no. 1, pp. 31–48, 2016.

[14] R. S. Cavalcanti, A. Moino Jr, G. C. Souza, and A. Arnosti, “Evaluation of the effect of pesticides on the development of the fungus Beauveria Bassiana (BALS),” VUILL. Arquivos do Instituto Biológico, vol. 69, no. 3, pp. 17–22, 2002.

[15] M. A. Altieri, “The ecological impacts of transgenic crops on agroecosystem health,” Ecosystem Health, vol. 6, no. 1, pp. 13–23, 2000.

[16] J. R. Lamichhane et al., "Integrated weed management systems with herbicide-tolerant crops in the European Union: Lessons learnt from home and abroad," Critical Reviews in Biotechnology, vol. 37, no. 4, pp. 459–475, 2017.

[17] S. E. Jacobsen, M. Sørensen, S. M. Pedersen, and J. Weiner, "Feeding the world: Genetically modified crops versus agricultural biodiversity," Agronomy for Sustainable Development, vol. 33, no. 4, pp. 651–662, 2013.

[18] H. Elver, "Promotion and protection of human rights: human rights questions, including alternative approaches for improving the effective enjoyment of human rights and fundamental freedoms," Interim report of the Special Rapporteur on the right to food, pp. 2–23, 2017.

[19] J. E. García G., "Cultivos genéticamente modificados: las promesas y las buenas intenciones no bastan," Revista de Biología Tropical, vol. 55, no. 2, pp. 347–364, 2007

[20] A. Ferrer and J. P. R. Cabral, "Epidemics due to pesticide contamination of food," Food Additives & Contaminants, vol. 6, no. 1, pp. 95–98, 1989.

[21] L. R. Goldman et al., "Pesticide food poisoning from contaminated watermelons in California, 1985," Archives of Environmental Health: An International Journal, vol. 45, no. 4, pp. 229–236, 1990.

[22] European Parliament, Directive 2009/128/EC of the European Parliament and the Council of 21 October 2009 establishing a framework for Community action to achieve the sustainable use of pesticides. 2009.

[23] M. D. A. Morales and S. D. P. González, "Resultados del seguimiento de plagas y enfermedades en cultivos de cereales en Tenerife. Campaña 2009. Especial referencia al trigo," 2009.

[24] R. Gabarra i Ambert, J. Moliner, and J. Arnó Satorra, "Control integrado de plagas en invernaderos de tomate temprano en la isla de Menorca," Boletín de sanidad vegetal. Plagas, vol. 20, no. 2, pp. 501–509, 1994.

[25] A. Fereres Castiel, "Impacto del cambio climático sobre los insectos vectores de patógenos de plantas," Phytoma España: La revista profesional de sanidad vegetal, vol. 300, pp. 105–109, 2018.

[26] E. R. Hunt Jr, S. I. Rondon, P. B. Hamm, R. W. Turner, A. E. Bruce, and J. J. Brungardt, "Insect detection and nitrogen management for irrigated potatoes using remote sensing from small unmanned aircraft systems," in Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping, 2016.

[27] B. Stumph et al., "Detecting invasive insects with unmanned aerial vehicles," in Proceedings - IEEE International Conference on Robotics and Automation, 2019, vol. 20, pp. 648–654.

[28] R. R. Shamshiri, I. A. Hameed, S. K. Balasundram, D. Ahmad, C. Weltzien, and M. Yamin, "Fundamental Research on Unmanned Aerial Vehicles to Support Precision Agriculture in Oil Palm Plantations," Agricultural Robots-Fundamentals and Application, pp. 91–116, 2018.

[29] P. Psirofonia, V. Samaritakis, P. Eliopoulos, and I. Potamitis, "Use of Unmanned Aerial Vehicles for Agricultural Applications with Emphasis on Crop Protection: Three Novel Case - studies," International Journal of Agricultural Science and Technology, vol. 5, no. 1, pp. 30–39, 2017.

[30] Y. Li, C. Xia, and J. Lee, "Detection of small-sized insect pest in greenhouses based on multifractal analysis," Optik, vol. 126, no. 19, pp. 2138–2143, 2015.

[31] R. Boll, C. Marchal, C. Poncet, and L. Lapchin, "Rapid Visual Estimates of Thrips (Thysanoptera: Thripidae) Densities on Cucumber and Rose Crops," Journal of Economic Entomology, vol. 100, no. 1, pp. 225–232, 2007.

[32] Y. Li, C. Xia, and J. Lee, "Vision-based pest detection and automatic spray of greenhouse plant," IEEE International Symposium on Industrial Electronics, no. ISIE, pp. 920–925, 2009.

[33] S. R. Huddar, S. Gowri, K. Keerthana, S. Vasanthi, and S. R. Rupanagudi, "Novel algorithm for segmentation and automatic identification of pests on plants using image processing," in 3rd International Conference on Computing, Communication and Networking Technologies, ICCCNT 2012, 2012, pp. 1–5.

[34] H. D. D. Nguyen and C. Nansen, "Hyperspectral remote sensing to detect leafminer-induced stress in bok choy and spinach according to fertilizer regime and timing," Pest Management Science, 2020.

[35] B. Lumbierres, M. Eizaguirre Altuna, R. Albajes García, and X. Pons, "Plagas de los espacios verdes urbanos: bases para su control integrado," Boletín de sanidad vegetal. Plagas, vol. 32, no. 3, pp. 373–384, 2006.

[36] R. Ebesu, "Integrated pest management for home gardens: insect identification and control," Insect Pests, vol. 13, pp. 1–11, 2003.

[37] M. L. Flint, "Pests of the garden and small farm: A grower's guide to using less pesticide," UCANR Publications, vol. 3332, 2018.

[38] Ministerio de Agricultura, Pesca y Alimentación, "Guías de gestión integrada de plagas," 2014. [Online]. Available: https://www.mapa.gob.es/es/agricultura/temas/sanidad-vegetal/productos-fitosanitarios/guias-gestion-plagas/

[39] Ministerio de la Presidencia, Agencia Estatal Boletín Oficial del Estado. Real Decreto 1311/2012, de 14 de septiembre, por el que se establece el marco de actuación para conseguir un uso sostenible de los productos fitosanitarios. 2012, pp. 11–13.

[40] J. Moral de la Vega, "Dificultad para diseñar una Gestión Integrada de Plagas adecuada a los complicados sistemas agrarios actuales," PHYTOMA España, vol. 237, pp. 90–92, 2012.

[41] Ministerio de Agricultura, Pesca y Alimentación, "SIAR. Sistema de información agraria para el regadío," 1998. [Online]. Available: http://eportal.miteco.gob.es/websiar/Inicio.aspx

[42] Ministerio de Política Territorial y Función Pública, "El sistema de información agroclimática para el regadío (SiAR)," datos.gob.es, 2019. [Online]. Available: https://datos.gob.es/es/blog/el-sistema-de-informacion-agroclimatica-para-el-regadio-siar

[43] D. C. Y. Vargas and C. E. P. Salvador, "Smart IoT gateway for heterogeneous devices interoperability," IEEE Latin America Transactions, vol. 14, no. 8, pp. 3900–3906, 2016.

[44] J. G. A. Barbedo, "Using digital image processing for counting whiteflies on soybean leaves," Journal of Asia-Pacific Entomology, vol. 17, no. 4, pp. 685–694, 2014.

[45] M. Qiao et al., "Density estimation of Bemisia tabaci (Hemiptera: Aleyrodidae) in a greenhouse using sticky traps in conjunction with an image processing system," Journal of Asia-Pacific Entomology, vol. 11, no. 1, pp. 25–29, 2008.

[46] C. Xia, T. S. Chon, Z. Ren, and J. M. Lee, "Automatic identification and counting of small size pests in greenhouse conditions with low computational cost," Ecological Informatics, vol. 29, no. P2, pp. 139–146, 2015.

[47] J. Cho et al., "Automatic identification of whiteflies, aphids and thrips in greenhouse based on image analysis," Iinternational Journal of Mathematics and Computers in Simulation, vol. 1, no. 1, pp. 46–53, 2007.

[48] L. O. Solis-Sánchez, J. J. García-Escalante, R. Castañeda-Miranda, I. Torres-Pacheco, and R. Guevara-González, "Machine vision algorithm for whiteflies (Bemisia Tabaci Genn) scouting under greenhouse environment," Journal of Applied Entomology, vol. 133, no. 7, pp. 546–552, 2009.

[49] L. O. Solis-Sánchez et al., "Scale invariant feature approach for insect monitoring," Computers and Electronics in Agriculture, vol. 75, no. 1, pp. 92–99, 2011.

[50] T. Liu, W. Chen, W. Wu, C. Sun, W. Guo, and X. Zhu, "Detection of aphids in wheat fields using a computer vision technique," Biosystems Engineering, vol. 141, pp. 82–93, 2016.

[51] R. G. M. Rupesh G. Mundada, "Detection and Classification of Pests in Greenhouse Using Image Processing," IOSR Journal of Electronics and Communication Engineering, vol. 5, no. 6, pp. 57–63, 2013.

[52] P. Boissard, V. Martin, and S. Moisan, "A cognitive vision approach to early pest detection in greenhouse crops," Computers and Electronics in Agriculture, vol. 62, no. 2, pp. 81–93, 2008.

[53] S. Poornima, S. Kavitha, S. Mohanavalli, and N. Sripriya, "Detection and classification of diseases in plants using image processing and machine learning techniques," in AIP Conference Proceedings, 2019, vol. 2095.

[54] J. Wang, Y. B. Chen, and J. P. Chanet, "An integrated survey in plant disease detection for precision agriculture using image processing and wireless multimedia sensor network," in Proceedings of the Internatinal Conference on Advanced in Computer, Electrical and Electronic Engineering ICACEEE, 2014, pp. 7–8.

[55] D. Jeric, A. Rustia, C. E. Lin, and T. Lin, "A Real-time Multi-class Insect Pest Identification Method using Cascaded Convolutional Neural Networks," in Proceedings of the 9th International Symposium on Machinery and Mechatronics for Agriculture and Biosystems Engineering ISMAB, 2018.

[56] V. Martin, S. Moisan, B. Paris, and O. Nicolas, "Towards a video camera network for early pest detection in greenhouses," in ENDURE International Conference, 2008, pp. 1–5.

[57] D. J. A. Rustia, C. E. Lin, J. Y. Chung, Y. J. Zhuang, J. C. Hsu, and T. Te Lin, "Application of an image and environmental sensor network for automated greenhouse insect pest monitoring," Journal of Asia-Pacific Entomology, vol. 23, no. 1, pp. 17–28, 2020.

[58] B. Vijayalakshmi, C. Ramkumar, S. Niveda, and S. C. Pandian, "Smart Pest Control System in Agriculture," in IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing, INCOS 2019, 2019, pp. 1–4.

[59] H. Darmono, R. H. Y. Perdana, and W. Puspitasari, "Observation of greenhouse condition based on wireless sensor networks," IOP Conference Series: Materials Science and Engineering, vol. 732, no. 1, 2020.

[60] S. Rodríguez, T. Gualotuña, and C. Grilo, "A System for the Monitoring and Predicting of Data in Precision Agriculture in a Rose Greenhouse Based on Wireless Sensor Networks," Procedia Computer Science, vol. 121, pp. 306–313, 2017.

[61] C. Calderon-Cordova et al., "Wireless sensor network for real-time monitoring of temperature, humidity and illuminance in an orchid greenhouse," Iberian Conference on Information Systems and Technologies, CISTI, vol. 2018-June, no. June, pp. 1–7, 2018.

[62] M. E. Bayrakdar, "Qualified Underground Wireless Sensor," IEEE Sensors Journal, vol. 19, no. 22, pp. 10892–10897, 2019.

[63] M. Erazo-Rodas et al., "Multiparametric monitoring in equatorian tomato greenhouses (I): Wireless sensor network benchmarking," Sensors (Switzerland), vol. 18, no. 8, pp. 1–22, 2018.

[64] M. Mancuso and F. Bustaffa, "A Wireless Sensors Network for monitoring environmental variables in a tomato greenhouse," IEEE International Workshop on Factory Communication Systems - Proceedings, WFCS, pp. 107–110, 2006.

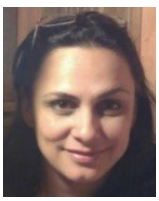
[65] J. Recasens and J. A. Conesa, Malas hierbas en plántula. Guía de identificación. Universitat de Lleida, 2009.

Miguel Ángel López

Has a degree in Engineering in Technical Engineering in Computer Systems from the University of Almería, graduates in Computer Engineering and Master in Softcomputing and Intelligent Systems from the University of Granada. Currently he is CTO. at Fidesol where performs diferent roles on the projects. His research focuses on distributed systems, management, integration and analysis of data, robotics, fuzzy logic systems, and the development of virtual reality environments for diferent purposes.

Juan Manuel Lombardo

Juan Manuel LombardoPhD in Computer Science from the Pontifical University of Salamanca, was graduated in Economics and Business Administration in the University of Granada, Spain, Diploma of Advanced Studies (DEA) in Economics from UNED, Research Sufficiency in Business Science from the Complutense University of Madrid and Diploma of Advanced Studies (DEA) in Sociology from the Pontifical University of Salamanca. He is CEO at Fidesol and Professor at Andalusia Business School. Dr. Lombardo is the author of numerous articles and research papers published in journals and books of national and international conferences. Visiting Professor at the Private Technical University of Loja (UTPL Ecuador), The National University of the Northeast (Argentina), University Francisco José de Caldas (Colombia), Catholic University of Colombia, Catholic University of Ibarra (Ecuador), University of Lisbon (Portugal) and National Engineering University (Peru). Member of the Knowledge Management committee of AEC (Spanish Association for Quality) and the Institute CICTES (Ibero-American Centre on Science, Technology and Society).



Mabel López

Has a degree of Computer Science Engineering. She is Knowledge Manager at Fidesol. Participates in the research and development strategy of this entity, technology transfer and analysis of technological trends, such as big data, internet of things, virtual reality, cognitive engines, machine learning, etc. Currently, she is involved in several R & D projects related to the mentioned technologies.

David Álvarez

David Alvarez received his Ph.D in Computer Science at the University of León, Spain in 2015, during which he worked on an Optical Caracter Recognition of handwritten documents. He then joined to the Research Group on Numerical Simulation and Scientific Calculus (SINUMCC) where he worked on Forest Fire Spread Simulation. He is now a member of the Research Group of Fundación del Software Libre (FIDESOL). His research interests include general computer vision, machine learning, image recognition and pattern recognition.

Susana Velasco

Has a Technical Engineer in Computer from the University of Granada. In the past, she worked in manufacturing, financial and service sector enterprises as software engineer and analyst programmer. Her research interests include quality assurance, software quality management systems, Ambient Intelligence (AmI) systems and devices, and new generation of ICT technologies.

Sara Terrón

Sara Terrón PhD in Business and Economics Studies from University of Granada, was graduated in Building Engineering and Master's Degree in Integral Safety in Building from Universities of Seville and Granada. Author of several papers, at Fidesol she currently focuses her research on the technological area.

Intelligent Detection and Recovery from Cyberattacks for Small and Medium-Sized Enterprises

Miguel Ángel López, Juan Manuel Lombardo*, Mabel López, Carmen María Alba, Susana Velasco, Manuel Alonso Braojos, Marta Fuentes-García

Fundación I+D del Software Libre (FIDESOL), Granada (Spain)

Received 16 May 2020 | Accepted 29 July 2020 | Published 7 August 2020

Abstract

Cyberattacks threaten continuously computer security in companies. These attacks evolve everyday, being more and more sophisticated and robust. In addition, they take advantage of security breaches in organizations and companies, both public and private. Small and Medium-sized Enterprises (SME), due to their structure and economic characteristics, are particularly damaged when a cyberattack takes place. Although organizations and companies put lots of efforts in implementing security solutions, they are not always effective. This is specially relevant for SMEs, which do not have enough economic resources to introduce such solutions. Thus, there is a need of providing SMEs with affordable, intelligent security systems with the ability of detecting and recovering from the most detrimental attacks. In this paper, we propose an intelligent cybersecurity platform, which has been designed with the objective of helping SMEs to make their systems and network more secure. The aim of this platform is to provide a solution optimizing detection and recovery from attacks. To do this, we propose the application of proactive security techniques in combination with both Machine Learning (ML) and blockchain. Our proposal is enclosed in the IASEC project, which allows providing security in each of the phases of an attack. Like this, we help SMEs in prevention, avoiding systems and network from being attacked; detection, identifying when there is something potentially harmful for the systems; containment, trying to stop the effects of an attack; and response, helping to recover the systems to a normal state.

Keywords

Attack Detection, Attack Recovery, Blockchain, Cybersecurity, Machine Learning, SME.

DOI: 10.9781/ijimai.2020.08.003

I. Introduction

CYBERCRIME is being increased to alarming levels nowadays, thus being already included in the national security and defense agenda. These crimes are a global epidemic that affect every computer system in the world [1]. The cybercriminal profile is not related to the expert and enthusiastic hacker that aims breaking security to test systems anymore [2]. Technically advanced countries and nations are now more involved in security incidents with different impact (due to either political or economical reasons) [3]. At the same time, criminal organizations tend to change their activity area in order to make their criminal practices more sophisticated. Cybercrime has become more professional, smart and stealthy [4]. This has lead to a change in the attacks, which are even more frequent in the last years.

Attacks aimed at exploiting vulnerabilities existing in information systems from critical infrastructures have been increased [5], [6], as well as from strategic areas such as energy or water supply, health, transports or finances [3]. In addition, Small and Medium-sized Enterprises (SMEs), due to their weakness and relevance in the activities and economy of a country, are also relevant targets. SMEs attacks aims to disturb or interrupt their basic structures, having a huge impact both in the entity [7] and the continuity of its services [8] that, sometimes, are essential. These attacks are well studied actions that imply a significant benefit with low risk for the criminals, due to its international nature, adaptability, mobility, and opacity [9].

Some of the most common attacks in 2018 were phishing, social engineering and data hijacking [10]. These attacks were performed achieving a 78% of effectiveness. Due to its fast propagation and effects in computer systems, which has even a more impact in SMEs, ransomware is one of the most important data hijacking attacks [11], [12]. This malicious software is really harmful [13], since it is diversified and it attempts to hide its actions and to maximize the benefits using advanced techniques [14]. When ransomware are activated, it is needed to switch off the systems and to activate all the security protocols for severe risks. The worst ransomware attack until the date was *WannaCry*, which took place in 2017. This attack affected many organizations and companies in 150 countries, having about 200,000 hosts affected [15]. The *Ryuk* attack, which was planned in a better way than *WannaCry*, took place more recently. *Ryuk* was designed to block as much number of systems as possible in a corporate environment [16]. On the other hand, the Covid-19 crisis has made the cybercriminals to be focused in health systems [17] and laboratories in order to worsen the consequences of the attacks [18].

Another dangerous example of stealing data is a bank Trojan that pursues the misappropriation of electronic bank accounts by means of collecting user credentials [19]. This attack has become more sophisticated since it appeared by 2004 for the first time. Checking the authenticity of a web page is not enough anymore: this malware fetches

* Corresponding author.
E-mail address: jmlombardo@fidesol.org

the introduced data, as well as the screen or data in the web page that is visited, making measures like virtual keyboards to be inefficient [20]. Distribution mechanisms in financial malware are better and better, which has serious effects in those entities that show a low defense against this type of attacks [21]. Financial malware increased 58% in 2019 in relation to the previous year, having more presence among threats, which is alarming [22].

Different malware families, such as ransomware, bank Trojans, and other attacks aiming to steal information, use the Domain Generation Algorithm (DGA) to generate many malicious domain names pseudo-randomly [23]. These domains can be used to compromise hosts, which makes it more difficult for the investigators to find the origin of the attack. Another attack that allows data stealing is SQL injection (SQLi), which introduces malicious code in a database by means of a web application, taking advantage of existing vulnerabilities in such database. Like this, the attacker can, for example, steal credentials for phishing the administrator identity and access, modify or delete information in the compromised database [24], even making it to be inaccessible. SQLi had high relevance during 2018 y 2019, being related to more than 72 % of attacks vectors to the web. From such vectors, 36% affected financial services directly [25]. Deny of Service (DoS) is also a dangerous attack against the availability of systems, which makes their legitimate users not being able to use compromised systems [26]. Distributed DoS (DDoS) evolved from the original DoS attack, having similar effects but applying different strategies. In a DDoS attack, the attacker usually builds an army (computers network, which is normally named botnet) by means of infecting hosts with malware (called zombies or bots). Bots can address the attack against a specific server, which ends suffering a heavy network traffic due to the overload [27]. The most advanced versions of DDoS are performed using Internet of Things (IoT) devices. This is the case of *Mirai*, *Brickerbot* or *Hajime* [28], which carry out DDoS attacks against low cost IoT devices that do not implement security measures. Like this, it is easier to control the infected devices, transforming them in an army that serves the hackers.

A. Mechanisms of Detection and Defense Against Attacks: the SME Problem

Both infrastructure and reputation turn damaged as a consequence of the aforementioned attacks. These damages can be even more severe for SMEs, having negative effects such as: reducing sells, losing clients [29], investors and number of employees, decreasing incomes, or even closing the company [30]. The number of cyberattacks increases everyday, which makes no organization to be free from suffering damages due to cybercrime. Furthermore, suffering a cyberattack in essential services provided by SMEs can produce a highly negative impact, yielding catastrophic effects if this happens in systems involved in critical infrastructures [31].

Although organizations invest in security mechanisms, most of these mechanisms are not effective due to attacks are more and more complex and evolve daily [32]. There is no doubt about sophistication and adaptability of cybercriminals to both environment and circumstances, since they study the most weak systems that are potential targets for the attacks [33]. This evolution is so fast that security teams can not predict the moment and target of an attack. Thus, it is essential to have a proactive security system allowing to detect threats and attacks timely in order to minimize damages. Companies are aware of this evolution and, for this reason, they invest in advanced security systems, such as Intrusion Detection Systems (IDSs), Security Information Event Management (SIEM) systems, Security Operations Centers (SOCs) or Managed Security Service Providers (MSSPs). These tools, by means of artificial intelligence, provide advanced threat and attack detection techniques and allow the automation of security processes [34]. Furthermore, companies create response plans according to systems and profiles in order to determine guidelines that need to be applied when a security incident is detected [35]. Yet, companies usually have a reduced economic capability to implement the aforementioned systems. Indeed, about 87% of companies do not have resources enough to acquire security systems [36]. This is the case of SMEs, which usually do not introduce the protection mechanisms needed. This makes them to be the main target for cybercriminals [37]. Furthermore, protecting new gadgets that are essential for working, such as smartphones or IoT devices is even more complicated [38]. For all these reasons, it is needed to develop a system with advanced features (similar to IDSs or SIEM systems), providing security against the most relevant attacks and being affordable for SMEs. This system should allow any organization to decrease both economic and social impact of suffering a cyberattack.

This paper shows the work carried out as a part of the IASEC project, which aims creating a researchers unit (*Unity of Innovation*) composed by workers both from Vector SF and Fidesol. The goal of IASEC is to perform research and development activities to build and optimize algorithms and tools, allowing to reach solutions that improve cybersecurity both in companies and other institutions. Initially, this project is being developed in a restricted environment for, later, being transferred to Andalusian SMEs for its evaluation under real conditions. The main objectives of IASEC are: *i)* providing resources to optimize detection and self-recovery of systems and services after suffering an attack, *ii)* creating a solution to allow detecting and dealing fake publications on the Internet, *iii)* protecting IoT devices and Industry 4.0 from the most relevant attacks for SMEs, and *iv)* detecting and avoiding fake news and hoaxes spreading. These objectives are tackled by combining both smart systems and blockchain. Like that, blockchain help us to improve the security systems by protecting data integrity in a secure and transparent way. Thus, the *Unity of Innovation* aims to be a reference point in relation to cybersecurity technology transference to Andalusian SMEs and institutions. In this work, we introduce an architecture for smart detection of the most important attacks for SMEs, addressing the first of the objectives in the IASEC project.

The rest of the paper is organized as follows. In Section II we review the literature related to cybersecurity incidents in SMEs. In addition, we explain some of the algorithms for detecting the most affecting attacks for such companies. In Section III, we introduce our proposal in the IASEC framework for solving the detection and recovery problems against the previously identified attacks. The platform developed for detecting these attacks, as well as the corresponding architecture, are also described in this section. Finally, in Section IV we present the main conclusions derived from this work.

II. Related Work

A. Cybersecurity Solutions for SMEs

Companies utilize different security tools with the aim of detecting and, sometimes, responding to security incidents. SIEM systems are one of the most extended tools. SIEM systems allow compliance of security regulations and managing events. These systems also allow event correlation as well as perform analysis of records and events from different data sources [39]. However, including a SIEM solution is really expensive and complex for SMEs [40]. IDSs are another security tool widely used. IDSs can be network-based (NIDS) and host-based (HIDS) [41]. NIDS monitor and analyze network traffic in real time, while HIDS analyze records, databases and other elements in a host to detect possible intrusions. Recently, researchers are focused in IDS development to achieve effective solutions against intrusions and attacks [42]. IDSs can also be grouped according to the type of detection technique. Thus, they can be signature-based [43] and model-based [44].

Different factors can be considered to choose a cybersecurity solution for an SME. For example, one can select indicators for the implementation of IDSs. Authors in [45] compare the main existing IDSs (*e.g. OSSEC*[1], *Snort*[2] or *Suricata*[3]), and then they normalize the obtained data, assigning quantitative values to each indicator (*e.g.* license type, type of IDS, operating system, and interface). Weighting values are fixed by each SME according to its needs. The results of this study show that the most accurate IDS for SMEs is Suricata [45]. Furthermore, authors in [46] analyze different solutions for protecting sensitive information in SMEs. As a result, they obtain a ten tools comparative, where IDSs are highlighted. Authors in [47] propose using model-based IDSs for SMEs. They use Machine Learning (ML) techniques for data collecting, testing and evaluation the proposal. Their main goal is to determine which is the most efficient algorithm for intrusion detection. To do this, they compare the following algorithms for supervised detection: C4.5 (Decision Tree), Bayesian Network, Random Forest, Support Vector Machines (SVM), and Artificial Neural Network (ANN). The study is performed by taking measures from different sampling data. Results show that C4.5 is the most precise among the studied algorithms [47].

Finally, another proposal is to build a solution focused in cybersecurity for Smart-Home or Smart-Office [48]. This work deals two research topics: data collecting from commercial or industrial IoT networks, and datasets exploitation for intrusion detection applying ML methods. For the last one, authors apply two variants of Long Short-Term Memory (LSTM), which is a type of neural network [48].

B. Detection and Response Algorithms for the Main SMEs Attacks

As we explained before, DoS (and DDoS), malware, or web-based attacks are some of the most extended security incidents [24]-[27]. Below we review some detection techniques for DDoS, SQLi and DGA, due to its impact for SMEs [26].

1. DDoS

Authors in [49] propose detecting DDoS attacks using Random Forests. The algorithm is validated using the *KDD´99 cup* dataset [50], which is labeled indicating whether exists an attack or not. The results of the study show that precision for attack detection is 94%, while 100% is reached for those that are attack-free [49]. Similarly, and using the same dataset, authors in [51] introduce a script to optimize the learning process. They start by selecting those features in the dataset that are more accurate for model building, thus reducing the training time. Then, they implement a Random Forest (motivated by results from a previous comparison) reaching 99,92% precision.

There exist solutions for recovery once a DDoS has taken place. For example, authors in [52] show that blockchain can be used to mitigate DoS attacks. To do this, they propose to create a smart contract and a blockchain infrastructure in *Ethereum*. When a server suffers a DoS attack, the system records in the smart contract those IP addresses that are involved in the attack, creating new blocks every 14 seconds. Thus, each user in this network has an updated list with malicious addresses in the interval, allowing the security people to take actions for attacks mitigation. This solution can be extended to DDoS attacks.

2. SQLi

Authors in [53] obtain a model for SQLi attacks detection by feature extraction from web traffic. These authors use a free dataset provided by the European Conference on Machine Learning and Principles and Practice of knowledge Discovery in Databases (ECML-PKDD) [54]. Expert knowledge allows selecting those features that help to detect patterns in web traffic related to SQLi attacks. Authors analyze and compare different detection algorithms using those features that have been selected. These algorithms are: Decision Stump, Naïves Bayes, Bayesian Network and Radial Basis Function (RBF) network, which is an ANN. The most efficient algorithm is Decision Stump [53]. Authors in [55] apply Naïve Bayes to classify SQL queries in malicious and legitimate. To do this, they take into account both grammar and SQL syntax, extracting features from language and defining rules. Another work that also apply feature extraction from SQL queries is [56]. Authors in this work train several classifiers, such as SVM, Ensemble Bagged Trees or Ensemble Boosted Trees. In this case, the best results are obtained for Decision Tree.

There exist solutions for SQLi attacks prevention and for system integrity preservation. For example, it is possible developing a blockchain system to avoid attacks against database management systems [57]. Authors in this work propose restricting access from nodes to the web server and the database. Access is filtered by using the blockchain, where the IP address used for accessing is recorded. Thus, only non-malicious IP addresses can access the server. Another work propose a framework that uses smart contracts from blockchain [58]. This framework has two components: the first one stores type of users and SQL queries, while the second one stores hash chains from queries that are allowed for each user. These chains are tokenized using the cryptographic function SHA256.

3. DGA

Authors in [23] propose classifying DGAs using LSTM. This proposal can be applied under real-time conditions, it is not based in features and allows classifying in families of DGA attacks [23], [59]. This type of neural networks are efficient for problems with sequential relationships, where previous states have effect in the current one [23]. LSTM is also applied in [60] for DGA classification. In this case, authors add a neuron with memory and the ability of discarding previous values that are far in time. DGA can be detected analyzing DNS traffic in pseudo-real time [61]. This work introduces an algorithm that is implemented using Aizoon Research for Advanced Malware Identification System (ARAMIS). The proposal filters non-resolved DNS requests (*UNRES*) and identifies those hosts showing the highest peaks for this value. When *UNRES* is increased suddenly, an analysis is performed, since it can be related to non-trusted domains trying a connection.

III. Detection and Response for Attacks Against SMEs. The IASEC Proposal

The IASEC project is composed of three main milestones. The first of them aims studying main threats for a company, as well as the stages of an attack from the point of view of defending the company. The goal is developing methods and tools for detection and self-recovery of systems and services after a cyberattack takes place. The second milestone aims tackling user identity management in the Internet by means of blockchain for digital identity certification. Finally, the last one aims obtaining the knowledge needed for detecting and managing fake information publication in the Internet. This should be performed by combining both ML and blockchain.

A. Steps for Defending an Attack

As explained during the Introduction, this paper is focused in detection and self-recovery of systems after an SME has suffered an attack. As a part of this first milestone, we have proposed a security platform for detection and self-recovery against cyberattacks. During the first stage of the project, we have designed and implemented the general architecture for the platform, as well as the detection (but not

[1] https://www.ossec.net/

[2] https://www.snort.org/

[3] https://suricata-ids.org/

the recovery) micro-services. Prior to this design and development, we have studied the steps for defending an attack, which are described in the next paragraphs from the perspective of the IASEC project.

1. Prevention

The goal of prevention is to avoid systems from being attacked and/or compromised, *i.e.* adopting required measures to make it more difficult carrying out a successful attack [62]. Prevention is essential, given that no company wants to deal with information stealing or denial of its services. These situations could involve serious economic consequences, as well as losing the reputation of the company [63]. Taking into account the relevance of preventing an attack, from IASEC we recommend that the implemented security system should include: access control, self-backup for critical files, self-update for security features, and black / white access lists. In addition, we believe that it is needed to install firewalls and anti-malware solutions to complement the aforementioned measures.

2. Detection

The goal of detection is to identify irregularities in systems [64]. This step is related to systems such as IDSs or SIEMs, which sometimes use ML to detect intrusions. In IASEC, we have studied some ML algorithms to detect the most affecting attacks for SMEs [23], [49], [51], [53], [55], [56], [60]. Then, we have selected the most accurate for each of the attacks, implementing an individual micro-service to detect each of them. These micro-services and the architecture proposed to implement our security platform are explained later in this section.

3. Containment

The goal of containment is to minimize the impact of having a cybersecurity incident in a company, avoiding its propagation and gaining time to build a recovery strategy [65]. Related actions can be disabling those accounts that have been compromised or isolating those hosts that have been infected [66], as well as making backups from hard-disks [67]. Another possibility is implementing fake systems that are similar to the real ones. These systems are used as traps, showing vulnerabilities that catch the attention from attackers, which makes the fake system to be attacked instead of the real one. Examples of this technique are honeypots or sandboxes [68]. From IASEC, we propose creating a virtual environment that should be designed to cheat the cybercriminals, emulating that it is possible to perform privilege escalation and steal user credentials. This solution is based in Deception Technology [69].

4. Recovery

The goal of recovery is to restore systems to a normal state after a security incident has taken place. To do this, it is needed to perform some actions after removing the threat [62]. Recovery, similarly to detection, is one of the most important stages in defense. Like this, recovery allows performing a fast restoration to the normal state of the organization, thus minimizing costs. From IASEC, we propose two solutions: *i)* creating a Security Incident Response Plan (SIRP) [35], and *ii)* implementing self-recovery measures. The former should include countermeasures in case of detecting any security breach. The latter is based in creating lists of malicious and legitimate IP addresses [52], [56]. Thus, we propose designing and developing a recovery system after attacks. This system should include a list containing malicious IP addresses that belong to the attacks previously detected using the micro-services. These IP addresses should be recorded in a blockchain network. Like this, it is possible to obtain a database updated with the malicious addresses for the server, allowing a faster recovery of the systems.

B. Architecture of the Platform

The proposed security platform is designed to be a system with an architecture organized in micro-services, which are deployed individually in dockers. Micro-services can communicate with the user using an API, as well as with other micro-services. The platform is implemented using different programming languages, depending on the needs, with higher priority for Python versus Java or .Net.

Fig. 1 shows the scheme for the architecture, the components and their relationships in the micro-services platform. This platform is composed of a front-end and a back-end. The front-end, which is represented in green color in the left part of Fig. 1, corresponds to a client application that allows users interacting with the security micro-services. The back-end is composed of micro-services and a relational database. Users can provide data, and those data that are generated by the platform are stored using a static storage. A non-relational database is also available for the algorithms, to support the big data processing. User-platform interaction is performed by the exposition of an *API REST*, a hub and a request balancer based in *Netflix OSS* (API Gateway, Service Mesh)[4]. Micro-services are run in a docker ecosystem (represented in the bottom of Fig. 1), which ensures running independence, high availability and scalability. Docker container receive load balancing, routing, and orchestration (docker swarm). The CORE module, which is represented in the right bottom corner in Fig. 1, includes multi-language services to provide the following functionalities: database connection, notification (*e.g.* e-mail), security, log recording, generation of files, managing the generated files or those that have been sent to the platform, and other transversal utilities to cybersecurity services.

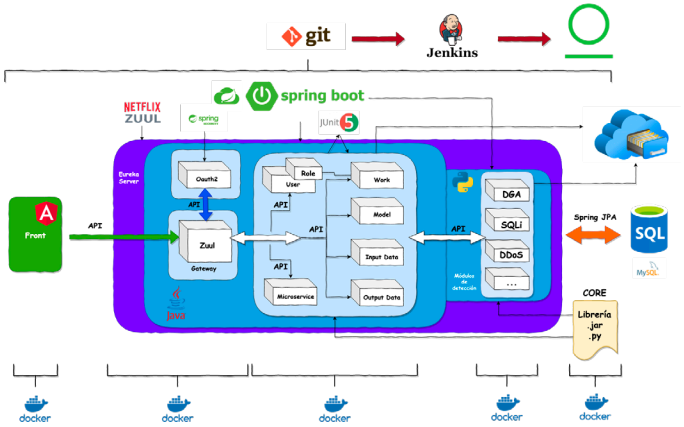


Fig. 1. Architecture for the IASEC micro-service platform.

Like this, our platform can be divided into two main parts:

1. **Part I**. This part manages user-related aspects of the platform: the database is created containing the tables that are needed for managing, like those related to users, micro-services and jobs (service requests). It is in charge of providing services to the user such as cybersecurity, data loading and downloading, authentication or job visualization, among others. Management is divided in *i)* web client (front-end), *ii)* back-end, and *iii)* load balancing (*Netflix* tools). The former ensures an agile interaction between user and platform. The second is composed by the management micro-services. Finally, the latter routes and balances algorithm requests for training and prediction, activating one instance of the docker where the requested algorithm is included.
2. **Part II**. This part involves the algorithms and the CORE module. The algorithms are run in instances of the docker, which will be stopped after each running due to the first phase of the project does

[4] https://netflix.github.io/

not allow real-time monitoring yet. Thus, intermediate information that is generated by the algorithm (*e.g.* support tables) is only available during the living time of the instance. In this first stage, micro-services allow loading a labeled dataset for the evaluation of the detection algorithms. The micro-service returns quantitative information related to the obtained prediction about detection. The micro-service output is saved into the static storage system.

C. Micro-services for Detection of Known Attacks

Attack detection micro-services are implemented using the python Micro-Framework, *Flask*[5]. Micro-services are scalable, and it is possible to connect them to a database or any other component. Each of them trains a model according to some features that has been previously defined and using a supervised ML algorithm. We have implemented an individual micro-service for each selected attack (recall DDoS, SQLi and DGA) under the hypothesis that it is more efficient to perform an independent detection (in terms of precision) than having a single algorithm to detect all the attacks. It is possible adding other modules in the future, like the self-recovery component.

Micro-services can access the storage system of the platform, where they can obtain files that have been uploaded by the user and write output files or saving models. Thus, micro-services input is the path for those files that are needed to process data, while output corresponds to a state indicating the result of the process (successful or not) and paths for the generated files. The attack detection micro-services need to perform the training using the input data. Training is composed of different steps: *1)* data pre-processing, *2)* model building (using the selected algorithm), and *3)* saving the model in a *pkl*[6] file, which is stored in the static storage system. Then, the generated model is used to carry out the prediction. This model is applied to new monitored data to detect whether there exist an attack or not. Please, recall that in this first stage of the project it is not possible to perform real-time monitoring, instead a labeled dataset can be loaded and the micro-service returns the detection rate and other relevant information about the dataset for positive detection. This output is also saved in the static storage system.

1. DDoS Detection

We have studied two different algorithms to implement the DDoS detection micro-service: Random Forest [49] and Decision Trees [51]. We have selected Decision Trees [49], since this algorithm obtain better results using the same dataset (*KDD´99 cup* [50]). For our testing, we have also used the *KDD´99 cup*, although it is possible to use any other dataset containing labels related to TCP connection, domain, and network traffic features [49].

2. SQLi Detection

We have selected Decision Stump [53] to develop the SQLi detection micro-service, since it obtain the best results among the studied algorithms (Naïve Bayes [55] and Decision Trees [56]). In this first stage, we have used the same dataset as in Reference [53] for our testing. Like for the DDoS micro-service, it is possible using any other dataset containing those features that are needed for model building.

3. DGA Detection

We have studied different proposals to implement DGA detection [23],[59]-[61]. We think that the best alternative is not only detecting DGA attacks, but also performing the classification among different families of DGA [23],[59]. Thus, we have decided implementing an LSTM neural network like in Reference [23], given its high precision both detecting DGA domains and classifying them. Our testing has been carried out using the domains database from Alexa[7] and a database that has been created in the IASEC framework, where the latest malicious domains detected by OSINT [59] are collected. Thus, we have trained the algorithm with two datasets: one containing legitimate domains and another with malicious domains. However, our micro-service accepts any dataset containing urls from domains.

D. Self-recovery

At the beginning of the section, we described the first milestone of the IASEC project, where one of the main goals is recovery for SMEs after suffering an attack. More precisely, one of these goals is self-recovery to protect the system integrity. Automation is usually related to design a tool being able of efficient self-recovery of the services in the compromised system. From IASEC, we propose implementing a module composed of micro-services being able to act in recovery tasks. Below we describe the proposed mechanisms for recovery after suffering any of the attacks that can be detected in our platform.

1. DDoS Recovery

We consider different mechanisms for resilience and recovery after suffering a DDoS attack. First, we propose creating a micro-service for recovery that uses blockchain to record IP addresses related to the attacks. The micro-service should discard those network packets containing malicious IP addresses by means of network traffic analysis, similar to the proposal in Reference [57]. This mechanisms can be improved creating black-lists and white-lists, also using blockchain [52]. Like this, the micro-service should prevent and recover services after the attack takes place. Additionally, if the server is being overloaded, we propose redirecting network traffic to alternative or backup servers if the origin IP addresses are not in the black-list. Thus, if the main server is suffering an attack but the IP addresses have not been identified as malicious, the server will maintain its services available.

2. SQLi Recovery

A possible solution to deal with an SQLi attack is locating those IP addresses that are injecting code and block them. To achieve this goal, we propose developing a micro-service for recovery. This micro-service should filter users requesting services taking into account their IP addresses. Queries that have been previously identified as malicious should be also denied. These IP addresses could be recorded using a smart contract in a blockchain. Like this, if a user is trying to access the system, a request should be performed to the blockchain network to allow or deny the access depending on the IP address in the origin [57]. Furthermore, this micro-service proposes a recovery mechanism for those tables that have been affected. This mechanism should check the writing history and restore to the last version before suffering the attack [70]. Like this, we propose a micro-service to avoid damages and allow recovering from an SQLi attack if finally it has been successful.

3. DGA Recovery

To deal with DGA attacks, we propose creating a black-list including those domains that have been identified as malicious using the detection micro-service. This black-list should be used to update the detection system. In the same way as for the rest of the attacks, the proposal is to develop a micro-service that should record the malicious domains in a blockchain network to provide integrity to the system. Additionally, the micro-service should be able of creating backups from critical files before the attack is completed. Like this, after suffering attacks such as ransomware (which is related to DGA), the micro-service should be able of recovering the system to a normal state, ensuring that no relevant information is damaged or corrupted.

Furthermore, we propose alerting the person in charge of security

[5] https://palletsprojects.com/p/flask/

[6] https://docs.python.org/3/library/pickle.html

[7] https://www.alexa.com/topsites

about each detected attack and the actions that have been carried out. Like this, he or she will be aware of the incident and will take needed actions. Finally, SME employees should be trained to avoid social engineering attacks and data hijacking, and should be also informed about the SIRP.

IV. Conclusion

In this work, the main cybersecurity problems for companies are analyzed, paying special attention to SMEs. In this sense, main available solutions to protect their infrastructures and systems are also studied. One of the main conclusions derived from this work is that, sometimes, early detection is even more important than prevention. Thus, detecting an attack in an early manner, allows the security team troubleshooting the incident properly.

The main cybersecurity problem affecting SMEs is that they do not have resources enough to set up efficient security systems, such as SIEMs. In this paper, we have proposed a solution considering each of the attack steps from the point of view of the protection of the company (prevention, detection, containment, and recovery). Our solution aims improving SME security, reducing both economical and social problems derived from suffering an attack. The proposal is part of the IASEC project. More precisely, we implement a security platform that provides different micro-services to detect DDoS, SQLi and DGA attacks. The proposed architecture is scalable, allowing to add new micro-services according to the needs of the SME. These micro-services can be both for detection and recovery. The results have been satisfactory for the first release of the platform, yielding a good basis for the next stages where it is expected that the attacks can be detected in real time.

As a future work, we propose developing a second release that provides micro-services for automatic recovery after suffering an attack. These micro-services will be blockchain-based in order to enhance the integrity of the system. In addition, new experiments will be performed using real network traffic, allowing to validate our current models with data from an SME. Finally, the IASEC project will continue, working in the certification of digital identities using blockchain and providing solutions to deal with fake information in the Internet.

Acknowledgment

The IASEC project has been funded by the IDEA Agency (Junta de Andalucía). Project code: 402C1800002.

References

[1] Kapersky, "Cyberthreat real-time map.Statistics", 2020. [Online]. Available: https://cybermap.kaspersky.com/stats/

[2] J. Salom, "El ciberespacio y el crimen organizado", Cuadernos de estrategia, no. 149, pp. 129-164, 2011.

[3] CCN-CERT, "Ciberamenazas y tendencias 2019", 2020. [Online]. Available:https://cutt.ly/JyxichC

[4] D. S. Wall, "Dis-organised Crime: Towards a distributed modelo of the organization of cybercrime", The European Review of Organised Crime, vol. 2, no. 2, pp. 71-90, 2015.

[5] L. Joyanes, "Introducción. Estado del arte de la ciberseguridad", Cuadernos de estrategia, no. 149, pp. 11-46, 2011.

[6] Council of the European Union, Council Directive 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection. 2008.

[7] M. S. Gordon, "Economic and National Security Effects of Cyber Attacks Against Small Business Communities", ProQuest Dissertations Publishing, 2018.

[8] S. Kamiya, J.-K. Kang, K. Jungmin, A. Milidonis and R. M. Stulz,"What is the impact of successful cyberattacks on target firms?", *NBER,* no. 24409, 2018. DOI:10.3386/w24409.

[9] Departamento de Seguridad nacional. Presidencia del gobierno, "Estrategia de seguridad nacional", 2013. [Online]. Available: https://cutt.ly/iyxinvl

[10] C. M. Arce, "Ciberseguridad y crímenes informáticos: el lado oscuro de la red", Revista Académica Arjé, vol. 2, no. 2, pp. 14-19, 2019.

[11] N. K. Popli and A. Girdhar, "Behavioural Analysis of Recent Ransomwares and Prediction of Future Attacks by Polymorphic and Metamorphic Ransomware", Computational Intelligence: Theories, Applications and Future Directions - Volume II. Advances in Intelligent Systems and Computing, vol 799, pp. 65–80, 2019. DOI:10.1007/978-981-13-1135-2_6.

[12] S. Bhattacharya and C. R. S. Kumar, "Ransomware: The CryptoVirus subverting cloud security", in 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), 2017, pp. 1-6.

[13] N. Scaife, H. Carter, P. Traynor and K. R. Butler, "Cryptolock (and drop it): stopping ransomware attacks on user data", in IEEE 36th international conference on distributed computing systems (ICDCS), 2016, pp. 303–312.

[14] M. Antonakakis, R. Perdisci, Y. Nadji, N. Vasiloglou, S. Abu-Nimeh, W. Lee and D. Dagon, "From Throw-Away Traffic to Bots: Detecting the Rise of DGA-Based Malware", in 21st USENIX Security Symposium, 2012, pp. 491-506.

[15] A. O'Dowd, "Major global cyber-attack hits NHS and delays treatment", BMJ, 2017. DOI:10.1136/bmj.j2357.

[16] J. Hernandez-Castro, A. Cartwright and E. Cartwright, "An economic analysis of ransomware and its welfare consequences", Royal Society Open Science, vol 7, 2020. DOI:10.1098/rsos.190023.

[17] M. V. Fontanilla, "Cybercrime pandemic", Eubios Journal of Asian and International Bioethics, vol. 30, no. 4, pp. 161-165, 2020.

[18] Dr. Rajib Subba, "Collective intelligence and international coordination: antidote for the novel biological zero-day exploit #COVID-19", Security Nexus Perspectives, 2020. [Online]. Available: https://apcss.org/wp-content/uploads/2020/04/Security-nexus-subba.pdf

[19] D. Kiwia, A. Dehghantanha, K.-K. R.Choo and J. Slaughter, "A cyber kill chain based taxonomy of banking Trojans for evolutionary computational intelligence", Journal of Computational Science, vol. 27, pp.394–409, 2018.

[20] P. Peris-Lopez and H. Martín, "Hardware Trojans against virtual keyboards on e-banking platforms – A proof of concept", AEU - International Journal of Electronics and Communications, vol. 76, pp.146–151,2 017.

[21] N. Tariq, "Impact of cyber-attacks on financial Institutions", Journal of Internet Banking and Commerce, vol. 23, no. 2, pp 1-11, 2018.

[22] V. Chebyshev, F. Sinitsyn, D. Parinov, B. Larin, O. Kupreev, E. Lopatin, "IT threat evolution Q1 2019. Statistics",2014. [Online]. Available: https://securelist.com/it-threat-evolution-q2-2019-statistics/92053/

[23] J. Woodbridge, H. S. Anderson, A. Ahuja and D. Grant, "Predicting Domain Generation Algorithms with Long Short-Term Memory Networks", Applied Sciences, vol.9 no. 20, 2019. DOI:10.3390/app9204205.

[24] W. G. J. Halfond, J. Viegas and A. Orso, "A Classification of SQL Injection Attacks and Countermeasures", in Symposium on Secure Software Engineering (ISSSE 2006), 2006.

[25] Akamai Research, "Financial Services - Hostile Takeover Attempts", State of the internet security, vol 6, no. 1, 2020.

[26] M. Jensen, N. Gruschka and N. Luttenberger, "The Impact of Flooding Attacks on Network-based Services" in Third International Conference on Availability, Reliability and Security, Barcelona, 2008, pp. 509-513.

[27] B. B. Gupta and O. P. Badve, "Taxonomy of DoS and DDoS attacks and desirable defense mechanism in a Cloud computing environment", Neural Computing and Applications, vol. 28, no.12, pp. 3655–3682, 2016.

[28] McAfee Labs, "Mirai, BrickerBot, Hajime Attack a Common IoT Weakness | McAfee Blogs," 2017. [Online]. Available: https://cutt.ly/UyxiQ9F [Accessed: Apr 17, 2020].

[29] S. Kamiya, J.-K. Kang, J. Kim, A. Milidonis and R. M. Stulz, "Risk management, firm reputation, and the impact of successful cyberattacks on target firms". Journal of Financial Economics, 2020. DOI:10.1016/j.jfineco.2019.05.019 2020.

[30] M.S. Gordon, "Economic and National Security Effects of Cyber Attacks Against Small Business Communities", ProQuest Dissertations Publishing, 2018. [Online]. Available: https://cutt.ly/WyxiRFu

[31] B. Genge, I. Kiss, and P. Haller, "A system dynamics approach for

assessing the impact of cyber attacks on critical infrastructures", International Journal of Critical Infrastructure Protection, vol 10, pp. 3–17. 2015. DOI:10.1016/j.ijcip.2015.04.001.

[32] M. A. Salitin and A. H. Zolait, "The role of User Entity Behavior Analytics to detect network attacks in real time" in 2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), 2018, pp. 1-5.

[33] A. Saravanan and S. S. Bama, "A Review on Cyber Security and the Fifth Generation Cyberattacks," Orient. J. Comput. Sci. Technol., vol. 12, no. 2, pp. 50–56, 2019.

[34] IBM, "Inteligencia artificial para una ciberseguridad más inteligente - España", 2020. [Online]. Available: https://www.ibm.com/es-es/security/artificial-intelligence. [Accessed: May 5, 2020].

[35] INCIBE, "¿Ya tienes tu Plan de Recuperación ante Desastres?," 2019. [Online]. Available: https://www.incibe.es/protege-tu-empresa/blog/tienes-tu-plan-recuperacion-desastres. [Accessed: May 5, 2020].

[36] Centro de estudios EY, "Global Information Security Survey, 9 de cada 10 empresas no cuentan con el presupuesto necesario para protegerse contra los ciberataques," 2019. [Online]. Available:https://cutt.ly/YyxoqEY [Accessed: Apr 17, 2020].

[37] Eleven Paths, "La importancia de la ciberseguridad en las Pymes" 2019. [Online]. Available: https://empresas.blogthinkbig.com/importancia-ciberseguridad-pymes/. [Accessed: Apr 29, 2020].

[38] J. Mesa, "IoT Dispositivos IoT dentro de la empresa : escenarios de ataque y protección", Red Seguridad, no.84, pp. 18–19, 2019.

[39] Gartner, "Security Information and Event Management (SIEM)",2020. [Online]. Available: https://cutt.ly/yyxotmu [Accessed: Apr 22, 2020].

[40] Gartner, "Magic Quadrant for Security Information and Event Management," 2020. [Online]. Available: https://cutt.ly/Zyxoiy8 [Accessed: Apr 22, 2020].

[41] J. Burton, I. Dubrawsky, V. Osipov, C. T. Baumrucker and M. Sweeney, "Introduction to Intrusion Detection Systems" in Guide to secure intrusion detection systems, Elsevier, 2003, pp. 1-38. [Online]. Available: https://cutt.ly/ayxooFL [Accessed: Apr 22, 2020].

[42] A. Boukhamla and J. Coronel, "Cicids2017 dataset: Performance improvements and validation as a robust intrusion detection system testbed", International Journal of Information and Computer Security, 2018.

[43] H. Holm, "Signature based intrusion detection for zero-day attacks: (not) a closed chapter?" in 2014 47th Hawaii International Conference on System Sciences, HICSS, IEEE Computer Society, 2014, pp. 4895-4904, 2014.

[44] V. Jyothsna, "A Review of Anomaly based Intrusion Detection Systems", International Journal of Computer Applications, no.28 , pp.26-35, 2011.

[45] N. D. Pantoja, S. A. Donado and K. M. Villalba, "Selección de indicadores para la implementación de un IDS en pymes", RISTI, no. E27, pp. 777–786, 2019.

[46] J. Waite, "Security Tools for the SMB and SME Segments", SANS Institute Information Security Reading Room, 2017.

[47] O. Elezaj, S. Y. Yayilgan, M. Abomhara, P. Yeng, and J. Ahmed, "Data-driven intrusion detection system for small and medium enterprises," in IEEE Int. Work. Comput. Aided Model. Des. Commun. Links Networks, CAMAD, pp. 1–7, 2019.

[48] N. Vakakis, O. Nikolis, D. Ioannidis, K. Votis, and D. Tzovaras, "Cybersecurity in SMEs: The smart-home/office use case", 2019 IEEE 24th Int. Work. Comput. Aided Model. Des. Commun. Links Networks, CAMAD, pp. 1–7, 2019.

[49] I. Moles, "Ancert: aplicación de técnicas de machine learning a la seguridad", Repositorio institucional (O2), 2018. [Online]. Available:http://hdl.handle.net/10609/88925

[50] Irvine, "KDD Cup 1999 Data," 1999. [Online]. Available:http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html. [Accessed: Apr 17, 2020].

[51] J. M. Rodriguez, "Aplicación de técnicas de Machine Learning a la detección de ataques", Repositorio institucional (O2), 2018.[Online]. Available: http://hdl.handle.net/10609/81126 [Accessed: Apr 17, 2020].

[52] J. Dheeraj and S. Gurubharan, "DDoS Mitigation Using Blockchain", Int. J. Res. Eng. Sci. Manag., vol. 1, no. 10, pp. 622–626, 2018.

[53] P. Aaby, "Evaluating Web App Datasets towards Detection of SQL Injection Attacks with Machine Learning Techniques", 2016. [Online]. Available: https://cutt.ly/fyxosIm [Accessed: Apr 17, 2020].

[54] C. R. Raïssi, J. Brissaud, G. Dray, P. Poncelet, M. Roche and M. Teisseire, "Web Analyzing Traffic Challenge: Description and Results", in The 18th european conference on machine learning and The 11th european conference on principles and practice of knowledge discovery in databases, 2007, pp.47-52.

[55] A. Joshi and V. Geetha, "SQL Injection detection using machine learning" in 2014 Int. Conf. Control. Instrumentation, Commun. Comput. Technol. ICCICCT, 2014, pp. 1111-1115.

[56] M. Hasan, Z. Balbahaith and M. Tarique, "Detection of SQL Injection Attacks: A Machine Learning Approach" in 2019 Int. Conf. Electr. Comput. Technol. Appl. ICECTA, 2019, pp.1-6.

[57] M. A. Mohd Yunus, M. Zainulariff Brohan, N. M. Nawi, E. S. Mat Surin, N. Azwani Md Najib and C. W. Liang, "Review of SQL Injection : Problems and Prevention", JOIV Int. J. Informatics Vis., vol. 2, no. 3–2, p. 215, 2018.

[58] M. Tmiezh, "A Framework for securing web applications against injection attacks using Blockchain technology", 2018, [Online]. Available: http://scholar.ppu.edu/handle/123456789/935. [Accessed: Apr 17, 2020].

[59] OSINT, "Feeds from Bambenek Consulting," 2019. [Online]. Available:https://osint.bambenekconsulting.com/feeds/.[Accessed: Apr 17, 2020].

[60] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory", Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[61] F. Bisio, S. Saeli, P. Lombardo, D. Bernardi, A. Perotti and D. Massa, "Real-time behavioral DGA detection through machine learning" in Proc. - Int. Carnahan Conf. Secur. Technol., 2017, pp. 1–6.

[62] Deloitte, "Pasos a seguir ante un ataque informático", 2020. [Online]. Available: https://www2.deloitte.com/es/es/pages/legal/articles/Pasos-a-seguir-ante-un-ataque-informatico.html. [Accessed: Apr 24,2020].

[63] Deloitte, "Los riesgos ocultos de un ciberataque," 2020. [Online]. Available: https://www2.deloitte.com/es/es/pages/governance-risk-and-compliance/articles/los-riesgos-ocultos-de-un-ciberataque.html. [Accessed: Apr 30,2020].

[64] Red Hat, "Manual de seguridad: Detección de intrusos," 2005. [Online]. Available: http://web.mit.edu/rhel-doc/4/RH-DOCS/rhel-sg-es-4/ch-detection.html. [Accessed: Apr 24, 2020].

[65] P. Cichonski, "Computer Security Incident Handling Guide : Recommendations of the National Institute of Standards and Technology", NIST Spec. Publ., vol. 800–61, p. 79, 2012.

[66] Bluegrass Group, "Containment," 2020. [Online]. Available: http://cybersecurityawareness.uk/recover/containment/. [Accessed: May 4,2020].

[67] Kaspersky, "Incident Response Guide Contents," 2017. [Online]. Available:https://cutt.ly/5yxojnV [Accessed: May 4,2020].

[68] Inforges, "CiberSOC: Gestión y monitorización de la seguridad informática en las empresas - Inforges", 2019. [Online]. Available: https://cutt.ly/ByxovlC[Accessed: Apr 30, 2020].

[69] Forcepoint, "What is Deception Technology? Deception Technology Defined and Explored" 2019. [Online]. Available: https://www.forcepoint.com/cyber-edu/deception-technology. [Accessed: Apr 24, 2020].

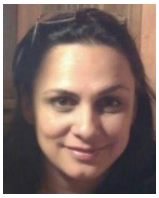
[70] K. Kurra, B. Panda, W. N. Li, and Y. Hu, "An agent based approach to perform damage assessment and recovery efficiently after a cyberattack to ensure E-government database security" in Proc. Annu. 2015 48th Hawaii Int. Conf. Syst. Sci., pp. 2272–2279, 2015.

Miguel Ángel López

Has a degree in Engineering in Technical Engineering in Computer Systems from the University of Almería, graduates in Computer Engineering and Master in Softcomputing and Intelligent Systems from the University of Granada. Currently he is CTO. at Fidesol where performs diferent roles on the projects. His research focuses on distributed systems, management, integration and analysis of data, robotics, fuzzy logic systems, and the development of virtual reality environments for diferent purposes.

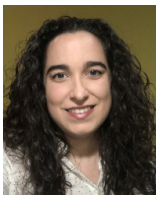
Juan Manuel Lombardo

PhD in Computer Science from the Pontifical University of Salamanca, was graduated in Economics and Business Administration in the University of Granada, Spain, Diploma of Advanced Studies (DEA) in Economics from UNED, Research Sufficiency in Business Science from the Complutense University of Madrid and Diploma of Advanced Studies (DEA) in Sociology from the Pontifical University of Salamanca. He is CEO at Fidesol and Professor at Andalusia Business School. Dr. Lombardo is the author of numerous articles and research papers published in journals and books of national and international conferences. Visiting Professor at the Private Technical University of Loja (UTPL Ecuador), The National University of the Northeast (Argentina), University Francisco José de Caldas (Colombia), Catholic University of Colombia, Catholic University of Ibarra (Ecuador), University of Lisbon (Portugal) and National Engineering University (Peru). Member of the Knowledge Management committee of AEC (Spanish Association for Quality) and the Institute CICTES (Ibero-American Centre on Science, Technology and Society).



Mabel López

Has a degree of Computer Science Engineering. She is Knowledge Manager at Fidesol. Participates in the research and development strategy of this entity, technology transfer and analysis of technological trends, such as big data, internet of things, virtual reality, cognitive engines, machine learning, etc. Currently, she is involved in several R & D projects related to the mentioned technologies.



Carmen María Alba

Degree in Information and Documentation at the University of Granada and web development technician specialized on Python. Currently, I am a document management in IASEC project and researcher in technologies as Big Data, Machine Learning and Blockchain.

Susana Velasco

Has a Technical Engineer in Computer from the University of Granada. In the past, she worked in manufacturing, financial and service sector enterprises as software engineer and analyst programmer. Her research interests include quality assurance, software quality management systems, Ambient Intelligence (AmI) systems and devices, and new generation of ICT technologies.

Manuel Alonso Braojos

Degree in Computer Science Engineering at the University of Granada. Currently he is a researcher at Fidesol where performs the role of fullstack software developer. He has participated in the development of various internet applications with different languages and technologies. His final degree work was an application for the representation of bibliometric data.

Marta Fuentes García

Holds a PhD in Information and Communication Technologies by University of Granada. She studied Computer Science and has a Master Degree in Software Development from the University of Granada. Her research has been mainly related to anomaly detection and diagnosis both in industrial processes and network traffic. Her PhD is focused in anomaly detection for network security using multivariate data analysis. She also has work experience in different companies as a programmer and, nowadays, she is part of the research team at Fidesol.

Smart Contracts with Blockchain in the Public Sector

Jenny Alexandra Triana Casallas[1], Juan Manuel Cueva Lovelle[2], José Ignacio Rodríguez Molano[3]

[1] Department of Computer Science, Universidad de Oviedo, Oviedo (Spain)
[2] Industrial Engineering, Universidad Distrital Francisco José de Caldas, Bogotá (Colombia)

Received 8 June 2020 | Accepted 16 July 2020 | Published 31 July 2020

Abstract

The appearance of so-called block chains or Blockchain with the promise of transforming trust and the way value is exchanged, joins the expansion of the technological capabilities of organizations to achieve higher levels of productivity and innovation. This is how Blockchain-based techniques are being applied to many fields, focusing in this article on the public sector, as a possible solution to the demands for transparency, participation and citizen cooperation that society demands; due to the possibility of disintermediation based on automated transactions and on the responsibility and security in the management of official blockchain records. This could obstruct corruption and make government services more transparent and efficient. Although, it investigates about applications in the public sector under the Blockchain system, such as transactions, agreements, property registries and innovations, developments and other assets; Special emphasis is placed on the possibility of implementing Smart Contracts (mechanisms that aim to eliminate intermediaries to simplify processes) in public procurement procedures, given that it is in this type of activity where high levels of corruption are generated. It is concluded then that Europe has the largest number of blockchain initiatives worldwide, while Latin America, except for the case of Peru, lacks this type of applications, being this continent exactly where there are the countries with the highest levels of corruption. It concludes with a recommendation to use blockchain along with smart contracts through platforms such as Ethereum or Lisk, mainly given its flexibility and current development on topics with similar functionalities.

Keywords

Blockchain, Smart Contract, Public Sector, Transparency, Corruption.

DOI: 10.9781/ijimai.2020.07.005

I. Introduction

The last decade has been characterized by a significant development of the Internet of Things (IoT), defined as the interconnection of objects connected to the Internet with the ability to communicate [1] and which is becoming part of organizations' technological capabilities and enable them to achieve increasingly accelerated levels of productivity and innovation [2]. In addition to the above, it is the artificial intelligence that is driving data mining [3]; Model-Driven Engineering (MDE) that emerges as the answer to the industrialization of software development because it provides better productivity and quality [4]; as well as the appearance of the so-called Blockchain is added with the promise of radically transforming the way in which value is exchanged [5].

The Blockchain emerged in 2008, as a proposal made by Satoshi Nakamoto [5], at a time when the deterioration of the economy, credit mistrust and the mortgage crisis, consumed the US market, and which would later spread to other countries [6]. Nakamoto's proposal aims to replace the centralized model with a decentralized one, where the decision-making power over the system is delegated directly to the users of the blockchain.

Blockchain-based techniques have become independent and evolved from Bitcoin, being applicable to many fields that go beyond currency, finance and markets [7], it is seen as a technical and economic innovation [8] [9] [10] particularly in the areas of government, health, science, literacy, culture and art [11], becoming a key factor in solving scalability, privacy and reliability problems directly related to the IoT paradigm [12].

A smart contract has de ability of executing and enforcing itself, autonomously and automatically, without intermediaries and is valid, without depending on authorities or third parties [8] by consensus of network users [13] [14]; eliminating bureaucracy given the decentralized, immutable and transparent nature of Blockchain technology.

Returning to Blockchain technology, and taking it to the field of government / State, within the framework of the society that demands transparency, participation and citizen cooperation; it is found that this technology potentially allows individuals and communities to redesign their relations in politics and society in general, with a process of large-scale disintermediation, based on automated transactions and responsibility and security in the management of official records [15] [16], which could favor the system of their states and eventually obstruct corruption and make government services more transparent and efficient [17], given that the blockchain offers a wide range of potential solutions in different governmental areas and that mainly benefit the citizen [18].

Therefore, Blockchain in public management could constitute decentralized solutions and consensus-driven public repositories, which may have a series of applications to make citizens less dependent on governments, but within a society that is ultimately based on State authority. Thus, providing the same services offered by the state and the corresponding public authorities (maintaining their validity), in a

* Corresponding author.
E-mail addresses: UO268296@uniovi.es (J. A. Triana Casallas), cueva@uniovi.es (J. M. Cueva Lovelle), jirodriguezm@udistrital.edu.co (J. I. Rodríguez Molano).

decentralized and efficient manner through Blockchain, does not mean dismissing the State, but promoting good government; this is "making better governments when all power is not concentrated in the hands of a few people" [17][19].

According to Preukschat [20], the public administration is in crisis because the way of understanding it is changing, and in order to adapt to these changes, it is necessary to promote new measures that guarantee an active and intelligent model, given that citizens demand more transparent, fast and efficient administration.

For this reason, integrating Blockchain in public management would suppose the solution of many of the problems that society frequently claims, and directly influencing organizations [9], at the same time that infrastructure cost savings could be incurred, given the relative simplicity in transactions using smart contracts [13].

This article provides a review of the level of progress in blockchain applications in public administration to find out what the future challenges and developments in this area could be.

II. Methodology

This work corresponds to the literature review related to the terms in both Spanish and English, in the advanced search engines and relevant databases, of: blockchain, smart contracts, open government, public administration, transparency and citizen participation.

After the selection of the conclusive terms for the research, they have been used to make queries in verified sources of information, such as forums recognized by the member states of the European Union, bibliographic and legal (legislative) databases, institutional and thematic repositories, platforms and indices on the level of national transparency:

1. EU Blockchain Observatory and Forum.
2. Bibliographic databases and academic and scientific repositories.
3. Information sources and database corresponding to current regulations and their respective modifications.
4. Transparency Portal of the Government of Spain, Transparency International - The Global Anti-Corruption Coalition, Transparency International Spain (TI - Spain) and Transparency Portal of the Generalitat Valenciana.
5. Transparency Indexes, the work analyzes the Transparency Index of the Autonomous Communities (INCAU) and the Transparency Index of the Municipalities (ITA).

Subsequently, a review was made of what is known, who and how blockchain has been studied and its repercussions in the different areas that can be introduced, as well as the origin and contribution of Smart contracts.

Knowing the bases, an analysis has been made of the cases in which blockchain-based applications or projects have been developed, of interest mainly to the public sector, showing them through a list of tables that allow analyzing the degree of implementation and viability following a pattern of own elaboration. Finally, a discussion was held on the proposed application, concluding observations, and analysis of future research.

III. Technical Context

First of all, it is important to review the definitions that exist of blockchain and Smart contracts, as well as their functions and applications in different contexts.

A. Blockchain

The blockchain is considered as a more recent technological revolution that can change the way of working of many industries; however, the greatest recognition has been achieved in the financial sector, due to the rise of bitcoin.

This term, popular for a few years, was born in 2009 associated with the aforementioned bitcoin, by Satoshi Nakamoto and is conceived as a decentralized, autonomous, auditable and reliable registry or database of transactions [18]; popularity also of the success of other cryptocurrencies [21]. Its origin is related to cryptography, linked to wars and power struggles between states, whose use intensified in the 1990s with cryptoanarchism, framed in the movement "cypherpunks" or activists who oppose surveillance of computer networks by states and evade censorship, defend the generalization of cryptography and technologies that improve privacy [20].

Thus, the Blockchain is a distributed peer-to-peer data structure in which untrusted members can interact with each other, verifiably without a trusted intermediary [22], this structure is shared and replicated between members of a network and in bitcoin, becoming the public and validated registry of all the transactions that have been executed [5] [23] [24] having the blockchain complete information about the balances of the initial block.

The potential of the blockchain is approached in a way that is accessible to different industries and sectors, including the internet of value model, different from the information internet.

The information available at the internet, allows its free circulation of information creating an infinity of business models; for its part, the internet of value is a tool to manage and share the value of digital assets or assets without the need to depend on a trusted central entity.

Blockchain became popular as the technology behind the Bitcoin cryptocurrency and as previously stated, it has emerged in other ways among applications such as smart contracts, which are scripts in the blockchain that allow automating multi-step processes, allowing self-fulfillment of digital contracts that are based on a counterfeit-proof consensus on contingent results, and financing through initial coin offerings [8].

Likewise, as with all new fashions, both technical and legal discrepancies and limitations arise when it comes to their large-scale application and it is essential to present and analyze them in order to carry out the review.

With the purpose of establishing an approach to the developments, applications and trends that originate in the context of BlockChain, a bibliographic review related to the subject is done in the first instance. The search continued to be delimited, including the criteria of blockchain and smart contracts in the areas of services, industry, logistics and public sector. As a result, the analysis proceeded with 21 documents [12], [25]-[44]. As a result of the review and analysis, Table I was obtained.

For the development of this research, it is important to note that, of the total number of markings (137), 13% (18) respond to problems associated with security, transparency and trust, 12% (16) correspond to developments supported by Smart contracts and 9% (13) are applications oriented to the public sector.

B. Smart Contracts

A very practical use of the blockchain is precisely smart contracts, which are tools that allow the automatic and independent execution of those terms of a contract that are programmable in relation to their functions through mathematical logic (IF + Then) and that make its clauses binding, unstoppable and automatic, and can be executed by external conditions. In addition, they guarantee the execution of a contract (neutrality principle), the delivery of digital goods and make the delivery of real goods and services more efficient.

Therefore, an smart contract can be understood as any contract

TABLE I. Topics and Applications Addressed with Blockchain in Services, Industry and Logistics

BlockChain in services, industry and logistics		
Grouping by topic	**Topics**	**Number of matching keywords**
Management, validation and registration of information	Registry maintenance	3
	Registration and protection of Industrial and Intellectual property	4
	BlockChain as a mechanism to provide Security, Transparency and Trust	18
BlockChain payment network	Cars with connectivity to make micro payments	1
	Air delivery drones that allow micro payments	2
BlockChain developments and tools	Platforms for the development of BlockChain Applications	1
	Ethereum blockchain and platform	5
	Smart contracts	16
	Digital Identity	7
Applications with decentralized features	DAOs (Decentralized Autonomous Organizations), DACs (Decentralized Autonomous Corporations) and DASs (Decentralized Autonomous Societies)	2
	Dapps (Distributed / Decentralized Applications)	8
BlockChain categories	Public BlockChains	9
	Private BlockChains	10
	Semi-private BlockChains	4
Traceability	Product life cycle	2
	Traceability and Logistics Systems (Supply Chain)	5
	Traceability in the automotive sector	1
	Transparency and Traceability in the Food Sector	3
	RFID	4
Applications in the area of technology and computing	Artificial Intelligence supported in BlockChain	1
	Applications that incorporate Internet of Things (IoT) and BlockChain	8
	Electronic Commerce	2
	Machine to Machine Interactions (M2M)	3
	CMfg (Cloud Manufacturing)	4
Environmental applications	Blockchain and Climate Change	1
BlockChain for business and business sectors	BlockChain and Insurance Companies	2
	Exchange of electricity through agreements supported in BlockChain	5
	Construction projects	1
	Exchange and Transaction of knowledge between teams and organizations	1
	Collaborative and Shared Economy	2
	Blockchain and Business Processes (BPM)	2

that executes by itself automatically without the measurement of third parties, but does not involve the use of artificial intelligence, they are written as computer programs or codes in which rules and consequences are defined and described, instead of being written in legal language on printed documents as they are traditionally known [27], since they act as binding agreements between two or more parties, and reside on the blockchain.

The term predates the blockchain with Nick Szaboo, who in 1994 was the first to use the term smart contract, defining it as a protocol for systematized transactions that execute the terms of a contract, mitigating the intervention of trustworthy third parties and avoiding malicious exceptions. They also inherit all the power that blockchain confers in terms of anonymity, security and decentralization [5].

IV. Blockchain and Smart Contracts Applications in the Public Sector

Public administration is one of the most relevant economic sectors because it is responsible for ensuring economic growth of a nation, as well as for establishing public policies that favor the social and economic well-being of citizens.

However, the current public administration does not provide comprehensive responses to society's demands, is not efficient, and is perceived as slow and bureaucratic [20]. So it is necessary to incorporate new perspectives to regain trust with more transparent, faster, more efficient and integrated models in the daily life of citizens, which also allows their participation and incidence, and in this sense, the blockchain system not only serves for companies seeking benefits, but also for public entities, both in the field of government, education, health, and in energy networks [28] , transport systems and social services, among others [29].

Taking into account that transparency was mentioned as a problem in the current public administration [18], it is crucial to change the behavior of an institution, you cannot force yourself to choose certain behaviors, but you can limit your decisions and actions.

In this sense, transcending the economic plane, the term "smart contracts" can be used as a pact of conditions in any field, which does not require intermediaries to validate or monitor compliance, which does not give rise to misleading interpretations and that they are completely transparent [30], because they execute themselves and are stored on a blockchain, which no one can control and that everyone can trust. In this sense, the blockchain system will allow these processes to be automated and will guarantee the integrity of their transactions, administrative concessions, records and important decisions, so that officials could not hide payments or official records or other manipulations from inside or outside and will favor greater control, traceability and transparency in the processes.

Currently, public administrations are going through a period of credibility crisis in many fields caused, in part, by the lack of innovative processes and the way of involving citizens in their decision-making [29]. The search for innovative resources to rethink communication and relationship with citizens must therefore be a constant in the management of Organizations and cities. The appearance of new currents of land management, such as those associated with the termination of Smart cities or the Internet of Things, cannot ignore the relationship with the citizen and the way in which he relates to and takes part in the own management of the territory [30].

Although in recent years, terms such as open government, transparency and electronic administration have been invoked as if they were modernizing public administration, it continues to lack the capacity to satisfactorily incorporate those instruments that would allow the most effective and efficient public activity. Therefore, an

administration incorporated into technological change in which the blockchain would make possible the aforementioned demands should be considered, by enabling citizens, companies and civil society organizations to access relevant information, improve public services and participate in decision-making more actively.

Among the benefits of blockchain is electronic voting, which decentralizes responsibility and disperses it among the participating nodes[12], which are the ones that achieve consensus on the data housed in the database by being based on centralized systems and governed by a single source.

Another application of blockchain in the public sector, is proposed by the Republic of Estonia, which has designed an e-government strategy based on interconnectivity and decentralization, openness and cyber security [29], which is why it is considered a leading country in digital government. and second with better indicators of social progress in terms of civil and political rights, along with Australia and the United Kingdom [29].

Estonia has recently launched its e-Residency program, in which anyone in the world can request a "transnational digital identity" and authentication to access secure services, as well as encrypt, verify and sign documents digitally. Currently the country is deploying a Blockchain system to streamline the sharing of various types of data between the Public Administration (legal, etc.), as well as its protection and security, and transparency.

Sweden, for its part, plans to build a blockchain-based property registry in search of security, modernity and transparency, thereby reducing the delay between signing a contract and registering it [27], the model has been exported to Georgia and Honduras.

In another case, in 2018 the signing of a Declaration for the establishment of a "European Blockchain Partnership" was signed, to address the digital transformation process that is looming for both public employees as well as citizens through the blockchain [31].

Already in Spain, the development of blockchain projects for the Public Administration is still in a premature phase, the uses of blockchain in Public Administrations are scarce, two of them are developed in the field of public procurement, one of the areas that, rightly, as suitable to "benefit" from this technology. This is the use of blockchain in the register of contractors, which was tendered by Basque Government´s Information Society (EJIE) in January 2018, from the Government of Aragon, to create a decentralized registry of public contract offers that allows a valuation afterwards automated offerings through smart contracts [32].

These initiatives in the form of contracting originate because the estimated amount of the weight of public procurement over GDP ranges from 18 to 20%, of which, the National Market and Competition Commission (CNMC) estimates that at least 48,000 million euros (4.5% of GDP) of the amount of the contracts must be associated with extra costs caused by the lack of transparency and competition in public procurement [32], for which the Council of the European Union issued a warning-recommendation to Spain to adopt measures aimed at reducing the deficit, expressing in this regard that in our country there is a "lack of sufficient a priori and a posteriori control mechanisms that hinder the correct and uniform application of public procurement legislation" [33]; as well as an absence of effective transparency.

Apart from the contracting projects mentioned in the previous paragraph, there is also the Project developed by the Alcobendas City Council: "Blockchain technology-based citizen participation voting system", the results of which were presented at the IV Congress of Smart Cities in 2018 [32].

On the other hand, in the same year, a Proposition of Law on regulation, taxation, communication of the legal use of cryptocurrencies and blockchain technology was approved by the Finance and Public Service Commission of the Congress of Deputies [34].

Table II summarizes the incursion of blockchain and Smart contracts in the public sector based on information from the European Union Observatory and Blockchain Forum [35].

The opportunities offered by the Blockchain are multiple, not only with a business vision. Regarding the public sector, some projects currently in the process of innovation are listed below:

1. Bit Nation

BitNation is a project based on smart contracts and Ethereum technology, and is defined as a "Decentralized Voluntary Nation without Borders" or digital nation. It is an open government, governance project that proposes solutions to have protected but demonstrable identity documentation, "public" coverage or insurance systems, management of "bitreputation" or reliability between commercial agents, and generation of procedures, such as birth certificates, among others [36].

2. D-Cent Project

D-Cent is a European-funded project that is in the research stage, to generate publicly owned "Citizen Participation Technologies", but seeking greater agility and public or diversified innovation. This project brings together various European initiatives, including those generated in Finland, Iceland and Spain (in the latter case, represented by projects promoted by the municipalities of Madrid and Barcelona).

One of the technological bases is the Blockchain, and thus seeks to generate technologies for the democratic management of Big Data (data generated by citizens and cities), to protect and ensure privacy and data protection with regulations, or manage spaces of digital public debate and deliberation [37].

3. DECODE Project

It is a European funding project that works on the idea of how citizens will be able to decide and manage their data in a scenario of greater transparency, automation and digitization of the data of cities and identities, as well as the economic impact that they could generate in these cities.

Barcelona and Amsterdam were the chosen cities to promote pilot projects. Specifically, in Barcelona it will revolve around the management of Open Democracy and the Internet of Things, while in Amsterdam it will focus on collaborative economies and the Internet of Things [38].

4. Property Registration

Blockchain allows the "tokenization" of assets, so that its transmission can be carried out with the confidence that the seller is who he says he is and can answer for the buyer, the payment is who he says he is and is the owner of the property that is transmitted and a Smart contract is in charge of automatically verifying all those circumstances and making payment for the property and registering it on behalf of the new owner [39].

In this sense, Germany, Dubai, South Korea, Ghana, Kenya, Singapore, Sweden, the United Kingdom, Brazil, Japan, the Indian state of Andhra Pradesh, the American state of Illinois, Ukraine, Croatia, Russia, etc., they are in the process or have already implemented property registration through Blockchain and try, to a greater or lesser extent, to solve similar problems such as fraud, corruption, transparency, the absence or multiplicity of data on the ground.

V. Results and Discussion

A. Potential of Blockchain in Public Procurement

The World Economic Forum estimates that the costs generated by

TABLE II. Blockchain and Smart Contracts Applications in the Public Sector

Country	Application	Status	Observations
Estonia	Online vote	Working	
Australia	Electronic Vote	Initiative	It would start with corporate and community elections before escalating to parliamentary elections.
United Kingdom	Online vote	Tests	
Sweden	Registration of property titles	Prototype	
Georgia	Registration of property titles	Prototype	
Honduras	Registration of property titles	Prototype	
Ghana	Registration of property titles	Prototype	
Russia	Registration of property titles	Prototype	The project aims to mitigate the possibility of corruption, while providing clients with secure and verifiable electronic receipts. It enables independent smart contract audits, as well as decentralized identity management, it has the potential to simplify the public registration process and ongoing maintenance through digital channels.
Switzerland	Online vote	Working	
Denmark	Online vote	Tests	
France	Online vote	Tests	
Holland	Online vote	Tests	
Australia	Digital identity	Initiative	It will allow people to verify their identity in a short time through a smartphone using biometric data.
China	Tax administration and electronic invoice issuance	Initiative	The Chinese government will use blockchain technology to organize and administer the system for collecting taxes and issuing electronic invoices. Reports indicate that China's existing tax system generates nearly two and a half trillion dollars in revenue, but officials believe there is a widespread tax evasion.
Dubai	Verification of electronic medical records between hospitals and clinics	Initiative	
Italy	Digital identity	Initiative	The pilot scheme will be implemented with the Estonian blockchain provider, and will include the existing manual process that will be replaced with a system that uses blockchain technology that can automatically transfer and verify patient records in seconds.
Vancouver	Public repository of verifiable claims on organizations.	No information	
Vancouver	Land registry	In verification	
United States	Electronic vote	In production	
United States	Elimination of paper records	No information	The Senate suggests that the use of a distributed ledger would eliminate the need for paper records and in-person updating of such data. Subsequently, the blockchain system would solve the existing data collection and retention problems in the state and create a more secure registry.
United States	Manage the identification of state residents, as well as tokenize assets in the public sector to improve efficiency and reduce rights fraud.	No information	Using a blockchain-based platform would allow state citizens to access and store all of their identifying information, such as taxes, voting and driver's licenses, etc., as decentralized nodes.
United States	Transfer of ownership		
	Digital identity		Birth records allow the state to issue a digital identity linked to the birth of a person that could be managed in a distributed ledger, adding attributes as the citizen interacts with different agencies throughout his life.
United States	Urban planning and public space (city planning and design)	Pilot	
	Citizen security		

corruption amount to more than 2.6 trillion dollars (more than 5% of the World Gross Product) [40], while in a report entitled "Myths and realities of governance and corruption", the World Bank has estimated that more than one trillion dollars per year is paid in bribes only [41]. This indicates that close to 2% of the World Gross Product ends up in the hands of corrupt government agents who intervene in the execution of various acts of the States and that, if the technology of smart contracts is implemented efficiently, they could be left without the power they use with illegitimate purposes for their own benefit or those of third parties, for example in public tenders that are automatically assigned to companies that have been the best bidders and not to those whose officials have offered some type of unofficial incentive or handouts, eliminating the possibility that There is an intermediary who can facilitate the completion of this payment or that the contractor collects without having executed the agreed work.

Using blockchain technology, each of the transactions can be traced to their origin, which contributes significantly to the prosecution of an eventual act of corruption. The differential that smart contracts provide over other solutions based on blockchain technology is in the self-execution of the instructions and the operations that they regulate, which generates an impossibility or, at least, an increase in the difficulty of executing acts of corruption.

Based on the above, it is important to state the issue of public procurement as one of the largest generators of corruption, given that it represents a substantial part of taxpayers' money worldwide, and remains the most vulnerable activity to waste, the fraud and corruption. Evidence of this is provided by data from the Open Government Partnership [42], which accounts for around 50% of the total spending of a typical government in low- and middle-income countries, and about 30% in high-income countries [43].

On average, 10-20% of procurement budgets can be wasted depending on the degree of corruption and waste and inefficiencies. Corruption distorts a fair adjudication system, limits the equality of opportunities between bidders, harming competition and consequently, decreasing the quality of public works, supplies and services, which also ends up undermining confidence in public institutions [43]

Recently the Communication COM (2017) 572 final from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, explicitly recognizes that the strictest provisions on the integrity and transparency of Directives are aimed at fighting corruption and fraud, and presents a public procurement strategy that establishes the general policy framework and defines clear priorities for improving procurement in practice and supporting investments within the EU, in the fight corruption in public procurement occupies a prevailing place [44].

On the other hand, the Organization for Economic Cooperation and Development (OECD) in its "Principles for integrity in Public Procurement" [45] maintains that integrity in public procurement is implemented in practice through four principles: good management, transparency, prevention of misconduct and accountability.

On the other hand, the Open Government Partnership (OGP) global report highlights that "*corruption in public procurement can reduce the value of contracts by up to 15% (depending on the estimate used). Open contracting - including the publication of contracts and citizen participation, monitoring and supervision - have shown that it has the potential to generate tax savings, reduce corruption and strengthen business participation*" [43], which demonstrates the great potential of block chain to improve indicators of fight against corruption worldwide.

In accordance with the aforementioned, the simple digitization of the contracting procedures from start to finish, and the establishment of publication of practically all the events of the procedure through applications in open and interoperable formats[46], consistent with the legislation of each country, and the multiplication of control mechanisms, introduce elements of transparency in the contracting process that by themselves reduce the possibilities of fraud, corruption and inefficiency. However, the characteristics of immutability, confidentiality, traceability and transparency of Blockchain together with the automation and disintermediation that Smart Contracts involve, makes it especially useful in the fight against corruption and fraud.

In addition, Botto and Castrovinci have pointed out as novelties that the use of Blockchain would report, in addition to the possibility of establishing a mechanism to control the integrity of the documentation and the process carried out by the bidding companies themselves, the reduction in the calendar of the procedures associated with tenders, given precisely because a Smart contract has already defined the execution rules [47]. In fact, if you plan to go beyond the award, Morris Gitonga points out that the use of Blockchain technology can prevent corruption in the management of awarded tenders insofar as all events are transparent and verifiable by each bidder [48].

In conclusion, given the characteristics of immutability, confidentiality, traceability and transparency of Blockchain together with the automation that Smart Contracts imply, it makes contracting procedures in the public sector the ideal field for its implementation, which would not eliminate corruption but if it would allow its early detection so that corrective and preventive measures can be taken.

Now for the development of this type of applications, a development platform is required that allows the incorporation of Blockchain and Smart contracts; within which, the University of Malaga in "On blockchain and its integration with IoT. Challenges and opportunities" [12] identifies various tools available for this type of development. Table III is a comparative table that evaluates this type of platform under four criteria: Blockchain type, consensus, cryptocurrency and smart contracts".

TABLE III. Blockchain Platforms for Creating Blockchain Applications (Taken From [12])

Platform	Blockchain	Consensus	Crypto currency	Smart contracts
Ethereum	Public and permission-based	PoS	Ether (ETH)	Yes
Hyperledger Fabric	Permission-based	PBTF/SIEVE	None	Yes
Multichain	Permission-based	PBTF	Multi-currency	Yes
Litecoin	Public	Scrypt	litecoins (LTC)	No
Lisk	Public and permission-based	DPoS	LSK	Yes
Quorum	Permission-based	Multiple	ETH	Yes
HDAC	Permission-based	ePoW,Trust-based	Multiasset	Yes

Based on Table III, the criteria to determine the most favorable platforms for the development of the application proposed here are evaluated, that is, they must allow the integration of Blockchain and Smart contracts as a minimum. Thus, Litecoin, Hyperledger Fabric, Multichain and Quorum are discarded, the first of them because it is not incorporating Smart contracts, and the others because they are limited to operating as managers of financial resources, under private

blockchain [49]. The rest meet the established criteria, however HDAC is ruled out, since it is a platform still under development [12]. Therefore Ethereum pioneered smart contracts and implementation in various areas [50] and Lisk, which enables the creation and implementation of decentralized applications, would be the recommended platforms to advance this solution [49].

Additionally, Ethereum and Lisk are determined as possible applications, because they are supported by public Blockchain technology, that is, it allows unrestricted access and the transparency parameter [51] prevails, allowing participation to be open, without loss of the attributes of security, transparency and traceability; this is just what it is required for the public sector.

The elements that make up the selected applications are built from the public BlockChain data structure and the Smart contracts that are derived from it, give the tool versatility without neglecting security, configuring a suitable space to be an axis of interaction, between the state, the bidders and the citizens interested in the contracting process, where each one can use and participate in the network, thus efficiently connecting to the various actors present, streamlining the stages of the process, having better access and data control, but allowing transparency in evaluation and adjudication.

B. Experiences of the Use of Blockchain in Public Procurement

There are several initiatives for the use of Blockchain in the public procurement that are presented below, however, no information on projects in operation was found except in the case of Peru:

- In Peru, the government agency for public procurement "Perú Compras" included the use of Blockchain in April 2018, to register purchase orders digitally. Since then, the country has registered nearly 50,000 purchase orders through its Electronic Catalogs platform [52]. Peru recognizes the application of Blockchain technology as an effective tool to provide transparency to the field of public procurement. Perú Compras operates through the Blockchain LAC-Chain network, a decentralized project of the Inter-American Development Bank (IDB) [53]. Since integration with the project, the objective is to promote the use of block technology among the countries of Latin America and the Caribbean.
- In Mexico, attendees at Talent Land 2018 were able to appreciate how a purchasing unit can make a call for tenders and how a company can apply to offer its products and services to the government, all through the blockchain, which makes transactions immutable and fully traceable [53]. This project currently has a design, a prototype in an alpha version with transactionality, it is expected to allow it to fulfill a period of maturity until reaching a beta version and then evaluate the possibility of applying it in a real case that goes along with Compranet, the Transactional system that allows public institutions in Mexico to carry out contracting procedures electronically, mixed or in person. The project has not been implemented due to regulatory problems that arise from introducing a technology into a government administrative process.
- In 2018, Canada successfully conducted the first ever use of public Blockchain technology (on Ethereum) in public contracts in order to enable transparent administration of government contracts [39].
- In the United States, the United States Government General Services Agency (GSA), through its Office of Emerging Citizen Technology, announced the launch of the United States Federal Blockchain Program with the objective that federal agencies and American companies can explore Blockchain technology [39].
- In Chile, in July 2018, a pilot test began for the use of Blockchain in public procurement, with the aim of improving Confidence indexes, transparency and less bureaucracy[54]. However, the results of the pilot test are not known and in September 2019, Chile established a cooperation agreement with Peru so that between both central Public Procurement they promote the exchange of successful experiences in state procurement processes and optimize electronic systems contracting in both countries.
- In Japan the application of Blockchain to Public Procurement is being tested. At the end of 2019, a meeting was held between Peru Purchases and the Korean contracting entity, in which the Asian representatives were interested in learning more about the application of technology in the public procurement sector [55].

C. Blockchain in the Tender and Award Procedure

The use of blockchain in public procurement should be oriented to respond to different models according to the legislation of the country where it is implemented, but there are common themes and activities that have been identified in the model for presentation and evaluation of offers proposed by Freya Sheer Hardwick, et al. [56], presented in Fig. 1.

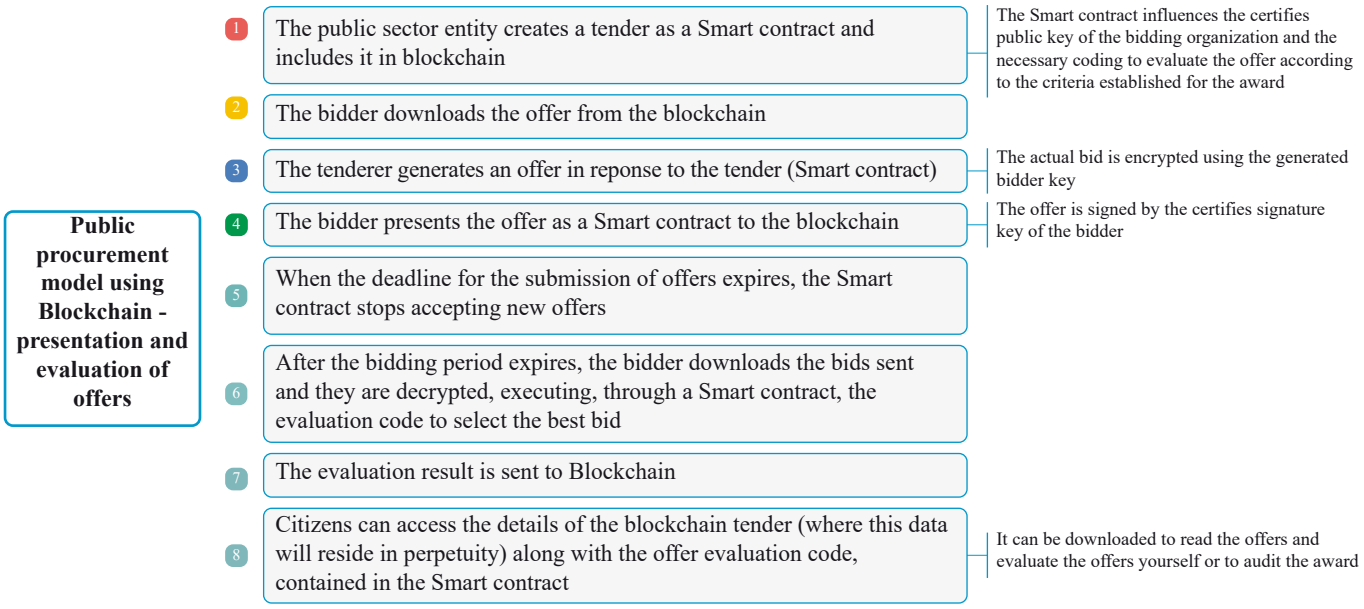


Fig. 1. Example of evaluation of bids under the smart contract approach (based on [56]).

D. Blockchain on the Confidentiality of the Documents of the Contractual Process

Blockchain technology can act as a guarantee of security and confidentiality in relation to the information that public sector entities place in the hands of the bidders in the bidding process, in which case the proof of the existence of consent for their access is marked and stored on Blockchain.

Obtaining consent must be a "block" before access to information classified as confidential by any subject outside the contracting authority. In fact, with the help of Blockchain cryptographic keys, no one will be able to access confidential information until they obtain the consent of the owner of said information. Each blockchain transaction can have an associated lock and the transactions can be pending and activated at a specific time in the agreed contract

Freya Sheer Hardwick, et al., cited above, point out a series of confidentiality and security requirements that in any case a tender system such as the one described should comply:

- Once bidders have uploaded their offer to the blockchain, they cannot modify it.
- The bidding organization cannot read the offer until the deadline expires.
- Bidders cannot change bids from another organization.
- Bidders cannot see who else made a bid.
- Blockchain network miners cannot affect the bidding process.

Therefore, the decentralized, transparent and secure properties of the Blockchain protocol can meet the conditions of public entities regarding the confidential information they handle, thus obtaining a more transparent and reliable process for its treatment.

E. Blockchain in the Procedure for Obtaining Guarantees

The presentation and return of guarantees, which are not applicable to all public procurement processes, refer to an additional novelty of complexity that can also be simplified and automated with the application of blockchain. The sequence is shown in Fig. 2.

VI. Conclusions

For the deployment of blockchain technology for the provision of services to the public administration, the following barriers should be overcome in the first instance:

At this time, there are areas of administrative action in which the use of blockchain is not allowed due to lack of legal coverage, so any development of technology of this type applied to solve any problem in the public sector, requires a review of regulations in the country that is intended to apply, among which is, the regulations associated with state contracting, transparency, good governance, budget, and those that regulate finance issues, whether at the national or territorial level (regional and local); the above so that the way to promote the use of blockchain is obtained, without later implementation having problems that prevent the development and maturity of the application.

In this case, the European Union, focuses on the regulation of transparency and cybernetics, while the lack of regulation in Latin America means that the foray into the issue has only just begun in countries like Peru.

On the other hand, when introducing blockchain technology in the administrative procedures of the public sector, it requires its detailed analysis and to be preceded by a review of the procedures on which they operate, a reflection on their need, and assess the possibilities of its simplification.

Also, the question of the infrastructure on which the blockchains will be deployed must be addressed, a question that cannot be addressed without taking into account the different administrative levels. It should be noted that this technology will bring multiple benefits, given that, unlike other technological fields, there are no obstacles to face [57], having as another point in favor that the implementation does not require large investments in terms of facilities and requires some personnel.

From a technical point of view, it is required that the solutions that can be developed in the framework of the use of blockchain in conjunction with Smart contracts for the contractual processes celebrated in the public sector, are implemented through platforms such as Ethereum or Lisk, given mainly its flexibility and current development in topics with similar functionalities.

However, it is important that the solution designed and implemented, supported in public blockchain, since not only in contracting but in most applications, it is sought that any citizen access the Blockchain and Smart Contracts without restrictions, based precisely on transparency and as a guarantee of the fight against corruption. Only in exceptional cases, such as voting platforms, is it necessary to shield the process with access restrictions, mainly due to identity, voting privacy and sensitive data management issues.

References

[1] C. Fang, X. Liu, P. M. Pardalos, and J. Pei, "Optimization for a three-stage production system in the Internet of Things: procurement, production and product recovery, and acquisition," *International Journal of Avanced Manufacturing Technology*, vol. 83, no. 5–8, pp. 689–710, 2016. [Online]. Available: https://search.proquest.com/openview/7dccddf108587ab6305afa24e74e89eb/1?pq-origsite=gscholar&cbl=2044010

[2] L. Antonio, S. Rubio, and A. B. Huertas, "Blockchain: La revolución de la confianza digital" (spanish), 2018. [Online]. Available: https://www.sic.gov.co/boletines-tecnologicos/blockchain-la-revolucion-de-la-confianza-digital. [Accessed: 07-Jun-2020].

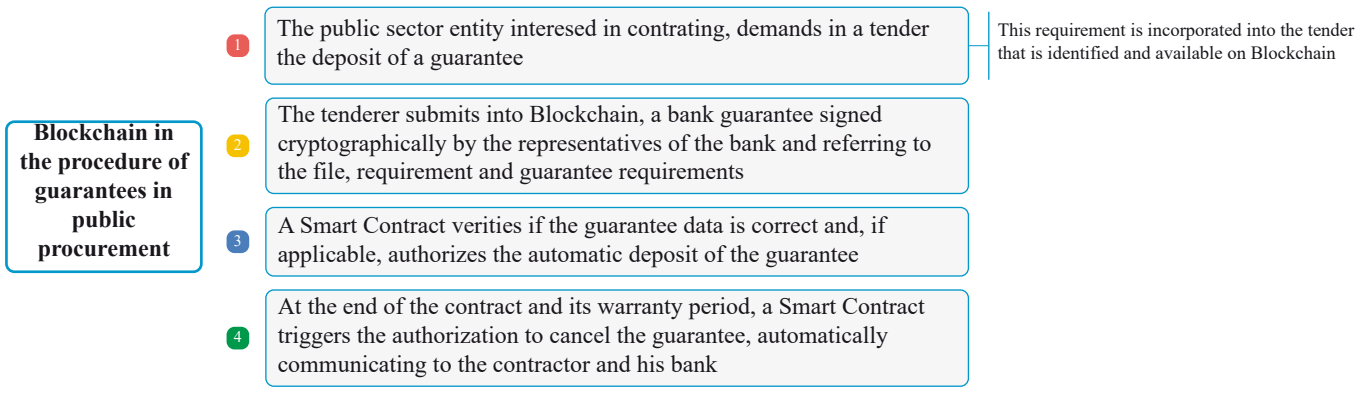


Fig. 2. Example of obtaining guarantees under the blockchain approach (based on [56])

[3] V. García-Díaz, J. P. Espada, R. G. Crespo, B. C. Pelayo G-Bustelo, and J. M. Cueva Lovelle, "An approach to improve the accuracy of probabilistic classifiers for decision support systems in sentiment analysis," *Applied Soft Computing Journal*, vol. 67, pp. 822–833, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1568494617303010

[4] V. García-Díaz, J. B. Tolosa, B. C. P. G-Bustelo, E. Palacios-González, Ó. Sanjuan-Martínez, and R. G. Crespo, "TALISMAN MDE framework: An architecture for intelligent model-driven engineering," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009, vol. 5518 LNCS, no. PART 2, pp. 299–306. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-02481-8_43

[5] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic cash system," *Bitcoin*, 2008. [Online]. Available: https://git.dhimmel.com/bitcoin-whitepaper/.

[6] M. J. C. Paul Vigna, *The Age of Cryptocurrency: How Bitcoin and Digital Money Are Challenging the ... - Paul Vigna, Michael J. Casey - Google Books*. 2015.

[7] R. Arjun and K. R. Suprabha, "Innovation and Challenges of Blockchain in Banking: A Scientometric View," *International Journal Interactive Multimedia and Artificial Intelligence*, 2020. [Online]. Available: https://www.ijimai.org/journal/bibcite/reference/2760

[8] L. W. Cong and Z. He, "Blockchain Disruption and Smart Contracts," *Rev. Financ. Stud.*, vol. 32, no. 5, pp. 1754–1797, 2019. [Online]. Available: https://academic.oup.com/rfs/article/32/5/1754/5427778

[9] D. Macrinici, C. Cartofeanu, and S. Gao, "Smart contract applications within blockchain technology: A systematic mapping study," *Telematics and Informatics*, vol. 35, no. 8. pp. 2337–2354, Dec-2018. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0736585318308013

[10] A. Savelyev, "Contract law 2.0: 'Smart' contracts as the beginning of the end of classic contract law," *Information and Communications Technology Law*, vol. 26, no. 2, pp. 116–134, 2017. [Online]. Available: https://wp.hse.ru/data/2016/12/14/1111743800/71LAW2016.pdf

[11] E. Portmann, "Rezension „Blockchain: Blueprint for a New Economy"," *HMD Praxis der Wirtschaftsinformatik*, vol. 55, no. 6, pp. 1362–1364, 2018. [Online]. Available: https://link.springer.com/article/10.1365/s40702-018-00468-4

[12] A. Reyna, C. Martín, J. Chen, E. Soler, and M. Díaz, "On blockchain and its integration with IoT. Challenges and opportunities," *Future Generation Computer Systems*, vol. 88, pp. 173–190, Nov. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X17329205

[13] M. Giancaspro, "Is a 'smart contract' really a smart idea? Insights from a legal perspective", *Computer Law and Security Review*, vol. 33, no. 6, pp. 825–835, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S026736491730167X#:~:text=Smart%20contracts%20are%20disintermediated%20and,technology%20for%20various%20commercial%20purposes.

[14] V. Shermin, "Disrupting governance with blockchains and smart contracts," *Strategic Change*, vol. 26, no. 5, pp. 499–509, 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/jsc.2150

[15] W. Reijers, F. O'Brolcháin, and P. Haynes, "Governance in Blockchain Technologies & Social Contract Theories," *Ledger*, vol. 1, pp. 134–151, 2016. [Online]. Available: http://www.ledgerjournal.org/ojs/index.php/ledger/article/view/62

[16] H. Hou, "The application of blockchain technology in E-government in China," in *2017 26th International Conference on Computer Communications and Networks, ICCCN 2017*, 2017, pp. 1–4. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8038519

[17] F. Casino, T. K. Dasaklis, and C. Patsakis, "A systematic literature review of blockchain-based applications: Current status, classification and open issues," *Telematics and Informatics*, vol. 36. pp. 55–81, Mar-2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0736585318306324

[18] P. García Mateo, "Blockchain aplicado al sector público (spanish)," Universidad Politécnica de Valencia, 2018. [Online]. Available: https://riunet.upv.es/handle/10251/111762

[19] A. Antonopoulos, "Thoughts on the future of money." [Online]. Available: https://www.meetup.com/es-ES/bitcoin-barcelona/events/229106313/. [Accessed: 06-Jun-2020].

[20] A. Preukschat Carlos Kuchkovsky *et al.*, *Blockchain: la revolución industrial de internet (spanish)*. Gestión 2000 (2017). ISBN: 978-958-42-6404-6.

[21] C. Q. Ardila, "La naturaleza económica del Bitcoin: un enfoque monetario" (spanish), 2018. [Online]. Available: https://www.mail-archive.com/cryptography@metzdowd.com/msg09959.html.

[22] K. Christidis and M. Devetsikiotis, "Blockchains and Smart Contracts for the Internet of Things," *IEEE Access*, vol. 4. pp. 2292–2303, 2016. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7467408

[23] C. Smith, "Blueprints for a new economy," *Nation*, 2011. [Online]. Available: https://www.thenation.com/article/archive/exchange-blueprints-new-economy/.

[24] T.-Y. Chung, I. Mashal, O. Alsaryrah, V. Huy, W.-H. Kuo, and D. P. Agrawal, "Social Web of Things: A Survey," in *2013 International Conference on Parallel and Distributed Systems*, 2013, pp. 570–575. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6808239

[25] M. Swan, *Blockchain: Blueprint for a new economy*. 2015. [Online]. Available: https://books.google.com/books/about/Blockchain.html?id=RHJmBgAAQBAJ&pgis=1

[26] J. Hwang *et al.*, "Energy Prosumer Business Model Using Blockchain System to Ensure Transparency and Safety," in *Energy Procedia*, 2017, vol. 141, pp. 194–198. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1876610217354437

[27] Ocariz Emiliano B., *Blockchain y smart contracts. La revolución de la confianza (spanish)*. 2019. RCLIA (2019). ISBN:978-84-948972-1-4

[28] D. Bodas-Sagi and J. Labeaga, "Using GDELT Data to Evaluate the Con dence on the Spanish Government Energy Policy," *International Journal of Interactive Multimedia and Artificial Intelligence*, 2016. [Online]. Available: https://www.ijimai.org/journal/bibcite/reference/2536.

[29] S. A. Tapscott, Don, Tapscott, Alex, *Blockchain Revolution*, 4th ed. Penguin USA (2019). ISBN: 978-84-234-2655-3

[30] J. A. Moralejo, "Blockchain en procesos de participación ciudadana" (spanish), in *Participación ciudadana:experiencias inspiradoras en España*, Centro de Estudios Políticos y Constitucionales. Mo de la Presidencia, Ed. España, 2018, pp. 147–158. [Online]. Available: http://www.gigapp.org/images/docus/Participacion%20ciudadana_11%20Arteaga.pdf

[31] M. N. Kamel Boulos *et al.*, "Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples.," *International Journal af Health Geographics*, vol. 10, no. 1, p. 67, 2011. ISBN: 1476-072X

[32] Observatorio de Contratación Pública (OBCP), "Blockchain, Administración y contratación pública" (Spanish) [Online]. Available: http://www.obcp.es/opiniones/blockchain-administracion-y-contratacion-publica. [Accessed: 22-Jun-2020].

[33] Comisión Nacional de los Mercados y la Competencia "E/CNMC/004/18 - Radiografía de los procedimientos de contratación pública en España" (Spanish). [Online]. Available: https://www.cnmc.es/expedientes/ecnmc00418. [Accessed: 22-Jun-2020].

[34] J. Berryhill, T. Bourgery, and A. Hanson, "Blockchains Unchained: Blockcahin technology and its use in the public sector," *OECD Working Paper on Public Governance*, no. 28, p. 53, 2018. [Online]. Available: https://oecd-opsi.org/wp-content/uploads/2018/06/Blockchains-Unchained-Guide.pdf

[35] "Initiative map | EUBlockchain." [Online]. Available: https://www.eublockchainforum.eu/initiative-map. [Accessed: 23-Jun-2020].

[36] S. T. Tempelhof, E. Teissonniere, J. F. Tempelhof, and D. Edwards, "Pangea Jurisdiction and Pangea Arbitration Token (PAT): The Internet of Sovereignty," no. April. pp. 1–42, 2017.

[37] "D-CENT" [Online]. Available: https://dcentproject.eu/. [Accessed: 24-Jun-2020].

[38] "DECODE" [Online]. Available: https://decodeproject.eu/. [Accessed: 24-Jun-2020].

[39] J. Berryhill, T. Bourgery, and A. Hanson, "Blockchains Unchained: Blockcahin technology and its use in the public sector," *OECD Work. Pap. Public Gov.*, no. 28, p. 53, 2018.

[40] "Smart contracts': 'blockchain' contra la corrupción | Compromiso Empresarial" (Spanish) [Online]. Available: https://www.compromisoempresarial.com/transparencia/2018/03/smart-contracts-blockchain-contra-la-corrupcion/. [Accessed: 22-Jun-2020].

[41] D. Kaufmann, A. Kraay, and M. Mastruzzi, "Measuring corruption : myths and realities." Africa Region Findings & Good Practice Infobriefs; no. 273. World Bank, Washington, DC. World Bank. pp. 1–5, 01-Apr-2007. [Online]. Available: https://openknowledge.worldbank.org/handle/10986/9576

[42] L. A. Rodriguez-Rojas, J. M. Cueva-Lovelle, G. M. Tarazona-Bermudez, and C. E. Montenegro-Marin, "Open Data as a key factor for developing expert systems: a perspective from Spain," *International Journal Interactive Multimedia Artificial Intelligence*, vol. 2, no. 2, p. 51, 2013. [Online]. Available: https://www.ijimai.org/journal/bibcite/reference/2414

[43] Open Government Partnership, "Open Government Partnership Global Report: Democracy beyond the ballot box," 2019. [Online]. Available: https://www.opengovpartnership.org/campaigns/global-report/#content.

[44] A. L. Comité, E. Y. Social, and E. Y. Al, "Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee and the Committee of the Regions," *European Commission*, 2018. [Online]. Available: https://ec.europa.eu/info/strategy/international-strategies/global-topics/sustainable-development-goals/eu-.

[45] Organisation for Economic Co-operation and Development (OECD), "OECD Recommendation of The Council on public integrity," 2017. [Online]. Available: https://www.oecd.org/gov/ethics/recomendacion-sobre-integridad-es.pdf.

[46] J. F. Herrera-cubides, P. A. Gaona-garcía, C. Montenegro-marín, and D. Cataño, "Security Aspects in Web of Data Based on Trust Principles. A brief of Literature Review," *International Journal Communication Networks and Information Security*, vol. 11(3), no. January 2020, pp. 365–379, 2019. [Online]. Available: https://www.ijcnis.org/index.php/ijcnis/article/view/4334/364

[47] "La blockchain negli appalti pubblici, come utilizzarla: i vantaggi | Agenda Digitale" (italian), [Online]. Available: https://www.agendadigitale.eu/procurement/la-blockchain-negli-appalti-pubblici-come-utilizzarla-i-vantaggi/. [Accessed: 29-May-2020].

[48] "Using Blockchain Technology to Eliminate Corruption in Developing Nations - coinweez." [Online]. Available: https://coinweez.com/using-blockchain-technology-eliminate-corruption-developing-nations/. [Accessed: 29-May-2020].

[49] C. Lin, D. He, X. Huang, K. K. R. Choo, and A. V. Vasilakos, "BSeIn: A blockchain-based secure mutual authentication with fine-grained access control system for industry 4.0," *Journal Network and Computer Applications*, vol. 116, pp. 42–52, 2018. [Online]. Available:https://www.sciencedirect.com/science/article/pii/S1084804518301619

[50] J. C. González, V. García-Díaz, E. R. Núñez-Valdez, A. G. Gómez, and R. G. Crespo, "Replacing email protocols with blockchain-based smart contracts," *Cluster Computing*, vol. 9, 2020. [Online]. Available: https://link.springer.com/article/10.1007%2Fs10586-020-03128-9

[51] I. Bashir, *Mastering Blockchain: Distributed ledger technology, decentralization and smart contracts explained*. 2018. [Online]. Available: https://books.google.com.co/books?hl=es&lr=&id=3ZlUDwAAQBAJ&oi=fnd&pg=PP1&dq=Mastering+Blockchain+Distributed+ledgers,+decentralization+and+smart+contracts+explained&ots=-4oZHGSKYK&sig=OyhPnlfY2XhUqmgR-F3EughnSSY#v=onepage&q=Mastering Blockchain Distribut

[52] "Perú Compras: Plataforma Digital única del Estado Peruano|Gobierno del Perú" (Spanish) [Online]. Available: https://www.gob.pe/institucion/perucompras/ [Accessed: 23-Jun-2020].

[53] "Convenio de cooperación entre ChileCompra y Perú Compras podría impulsar adopción Blockchain en contratación pública - DiarioBitcoin." [Online]. Available: https://www.diariobitcoin.com/convenio-de-cooperacion-entre-chilecompra-y-peru-compras-podria-impulsar-adopcion-blockchain-en-contratacion-publica/. [Accessed: 24-Jun-2020].

[54] Organisation for Economic Co-operation and Development (OECD), "Council Report on the implementation of the recommendation of the Council on public procurement JT03449826," 2019. [Online]. Available: https://one.oecd.org/document/C(2019)94/FINAL/en/pdf

[55] M. Gonzalez, "Blockchain in Japan", EU-Japan Centre for Industrial Cooperation. 2018. [Online]. Available: https://www.eu-japan.eu/sites/default/files/publications/docs/blockchaininjapan-martagonzalez.pdf

[56] F. S. Hardwick, R. N. Akram, and K. Markantonakis, "Fair and Transparent Blockchain Based Tendering Framework - A Step Towards Open Governance," in *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, 2018, pp. 1342–1347. [Online]. Available: https://ieeexplore.ieee.org/document/8456054

[57] Cueva-Lovelle, "Internet de las cosas e industria 4.0" (spanish), Oviedo University (2018). [Online]. Available: http://di002.edv.uniovi.es/~cueva/conferencias/2018-09-13-LECCION-INAUGURAL.pdf. [Accessed: 07-Jun-2020].

Jenny Alexandra Triana Casallas

Industrial Engineer, Francisco José de Caldas District University, Colombia, 2009. PhD Student in Computer Science, University of Oviedo, Spain. Master in Organization Administration, National Open and Distance University, Colombia, 2014. Specialist in Hygiene and Occupational Health, Francisco José de Caldas District University, Colombia, 2010. Public Management Specialist, National Open and Distance University, Colombia, 2018. Associate Professor at Manuela Beltrán University, Bogotá- Colombia. Research interests: Blockchain, Smart contracts.

Juan Manuel Cueva Lovelle

Professor at the University of Languages and Computer Systems at the University of Oviedo (Spain). Mining Engineer (1983). Doctor from the Polytechnic University of Madrid (1990). He was Director of the University School of Technical Engineering in Informatics of Oviedo (University of Oviedo) from July-1996 to July-2004. He was Director of the Department of Informatics at the University of Oviedo from 2008 to 2016. He was coordinator of the Master in Web Engineering at the University of Oviedo from 2005 to 2016. His research areas are Blockchain, Internet of Things (IoT) Industry 4.0, Smart Mining, Drone Applications, Language Processors, Human-Computer Interaction, Model Driven Engineering and Web Engineering. He has directed more than 25 Research projects, more than 100 contracts with companies and 35 doctoral theses. He is the author of more than 200 books, articles and communications to congresses.

José Ignacio Rodríguez Molano

Doctor in Computer Engineering from Oviedo University (2017). Master in Web Site Management and Engineering at the International University of the Rioja - UNIR (2012). Master in Information Science at the Universidad Distrital Francisco José de Caldas (2012). Specialization in Geographic Information Systems the Universidad Distrital Francisco José de Caldas (2002). Industrial Engineer the Universidad Distrital Francisco José de Caldas (1998). Associate professor, attached to Engineering Faculty of Universidad Distrital Francisco José de Caldas", since 2006. Coordinator Master in Industrial Engineering (February 2018 –Current). Coordinator of the industrial engineering program (December 2013 – February 2017). Research interests: Blockchain, Smart contracts, Smart Cities, Industry 4.0, Internet of Things (IoT).

Blockchain-Enabled Platforms: Challenges and Recommendations

M. Inmaculada García Sáez*

Warwick Business School (United Kingdom)

Received 23 March 2020 | Accepted 25 April 2020 | Published 26 August 2020

Abstract

Not even a tenth of blockchain-enabled platforms survive their first anniversary. The volatility of crypto-markets has brought negative attention and led some to question the applicability of blockchain technology. This paper argues that the challenges for startups and incumbents behind these platforms are numerous, and that the speculative bubble around cryptocurrencies is only one of them. Blockchain still needs to demonstrate fully its disruptive potential and so far, entrepreneurs have not managed to significantly impact incumbents' market shares. This transitory period requires incumbents to let go of traditional control mech-anisms, and startups to scale down their global decentralised hopes. Indeed, whilst the technology can indeed scale fast, starting in a controlled market and managing growth is a counterintuitive but essential strategy for blockchain-enabled platforms to implement. Given the diverging nature of the technology, at present at least, the combined shortage of skills in blockchain and security, and the trust blockchain is built on, rushing to the global market is high risk. Nonetheless, given the potential returns, the risk appetite is high and both entrepreneurs and corporate executives share unrealistic expectations about a technology they cannot fully understand since it has not yet converged. In light of the above, this article identifies the main challenges faced when building blockchain-enabled platforms and provides recommendations for startups and incumbents to overcome these. In order to reach these conclusions, the information obtained from twenty semi-structured interviews with leading actors in the field has been fundamental.

Keywords

Adoption, Blockchain, Collaborative Networks Disruption, Value Proposition.

DOI: 10.9781/ijimai.2020.08.005

I. Introduction

In the last decade, the unprecedented rate of disruptive digital innova-tion has raised concerns about the sustainability of today's firms, jobs and business models. It has also sparked a wide range of new oppor-tunities [1]. The emergence of platform businesses and the sharing economy has rapidly transformed industries, leveraging innovative business models that changed consumer preferences, from Airbnb to Uber.

The emergence of the sharing economy, the rise of platform businesses and the birth of bitcoin are closely intertwined with the global financial crisis in 2008 [1]-[3]. These developments offered a shift from traditional centralised systems, enabling peer-to-peer value exchanges with a shared interest to co-create alternative systems. In this context, the sharing economy had one major obstacle to disintermediate the access to resources that blockchain solves: trust.

"The one thing that's missing, but that will soon be developed, it's a reliable e-cash. A method where buying on the Internet you can transfer funds from A to B, without A knowing B or B knowing A" [4].

What blockchain promises, and what fuelled its popularity, is no less than the technological backbone of the 21st century's renaissance of the social commons [5]: increase in transparency and horizontal cooperation, fair distribution of wealth and decrease in top-down governance, corruption, censorship or coercion [6]-[8].

Tetris, the popular game from the 90s, is a prime example of the world then: top-down, hierarchical, rigid rules, impossible to beat. Minecraft, a popular game nowadays, is all the contrary: bottom-up, players fall under no authority, do what they want and win all the time. With the explosion of social media and increased data leaks on corruption in centralised systems, social unrest is growing all over the world. In other words, people have a Minecraft mindset, but this is still very much a Tetris world [9]. Blockchain is the technology that makes a Minecraft world possible. Blockchain-enabled platforms are where this world will be created.

Indeed, by 2025, ten percent of the global gross domestic product is expected to be stored on blockchain technology [10]. And blockchain-enabled platforms will go beyond the main characteristics of the technology, from transparency to immutability, to actually reinvent the way we do business, the way we capture value and hopefully solve some of the world's societal challenges; in short, it can bring about a post-capitalist world.

However, the ability for blockchain to unleash the full potential of the Fourth Industrial Revolution, to design and govern more inclusive platforms that could create enough wealth to close the economic divide, is undermined by numerous adoption barriers [11]. In 2018, the average lifespan of a blockchain project was 1 year and two months. 92 percent of blockchain projects simply fail [12]. Given the potential of the technology, this fact only justifies the need for further research.

* Corresponding author.
E-mail address: inmaculadagarcia@gmail.com

The objectives of this paper are to analyse the challenges that blockchain-enabled platforms face and provide recommendations on how to overcome these.

II. Literature Review

A. Introduction: the Birth of Blockchain

"Getting a global society to agree that something has value and can be used as a currency without government support and without a physical form is one of the most significant accomplishments in monetary history" [6].

The Bitcoin white paper introduced a peer-to-peer electronic cash system [13] outlining the basic architecture of a public, distributed and immutable electronic ledger, better known as blockchain. Using cryptography and an innovative consensus mechanism called Proof-of-Work (PoW), it allows for trustless and decentralised transactions. Blockchain technology, or simply blockchain, refers to the derivatives of Bitcoin's blockchain.

Blockchain is often described as a global, shared and distributed database [14]-[15] however unlike in traditional databases, records on the blockchain are immutable [16] and allow for users to retain the ownership of their assets. Blockchain was the first implementation of a new type of databases, commonly referred to as Distributed Ledger Technologies (DLTs), although most practitioners continue to use the original term, blockchain, and so will this paper.

A review of the existing literature follows in order to understand the current and future challenges faced by blockchain-enabled platforms from three perspectives: innovation, platforms and technology.

B. Innovation Perspective

1. Blockchain as an Innovation

"The technology most likely to change the next decade of business is not the social web, big data, the cloud, robotics, or even artificial intelligence. It's the blockchain, the technology behind digital currencies like bitcoin" [17].

The difficulty to define the type of innovation that blockchain represents stems from the fact that there are multiple and often contradicting perceptions about its potential (Fig. 1).

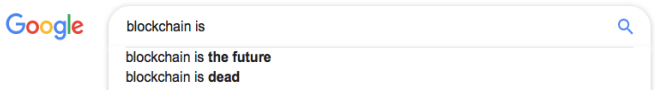


Fig. 1. Contradictory perceptions of blockchain potential.
Source: author based on Google search.

Among those who indeed view it as innovative, two main groups stand out. The first conceives blockchain as a foundational technology [18] that enables social and economic progress, but not as a disruptive technology that could represent a threat for incumbents [19].

The second group advocates that blockchain is more than just a foundational technology to be used by incumbents: its decentralization potential is disruptive [20]-[22]. Coase's Theory of the Firm, despite dating from 1937, is particularly relevant in this context as it suggests that the economic system self-organises the distribution of resources with no need of central authority [23]-[24]. Certain firms' activities would therefore no longer be justified, considering blockchain's ability to reduce transaction costs and networking costs [8], [23], [25], [26]. For example, blockchain allows for instant and frictionless international payments seriously undermining existing brokerage functions.

Christensen's theory of disruptive innovation explores how startups can unseat established incumbents by offering products appealing to a low-end or new market that is often overlooked by incumbents, as these focused on satisfying most profitable segments. When challengers improve their offering and move to mainstream customers, incumbents face the so-called innovator's dilemma of competing with the challenger and risk to cannibalise profits from higher-priced models [27], [28]. Blockchain offers new value creation opportunities that were not economically viable until now. For example, blockchain applications for remittance services [1], an essential part for the financial inclusion of those unbanked, is a segment that until now had not been of interest to established financial institutions.

Nonetheless, it should be noted that disruption refers to a process that can take decades, explaining why incumbents often overlook the disrupters. The ability for blockchain startups to challenge incumbents will depend on their ability to move from inferior to high-quality services in the eyes of incumbents' mainstream customers [28].

2. Blockchain Innovation Adoption

The adoption of foundational technologies, such as internet and blockchain, often occurs in phases: single use, localised use, substitution and transformation [19]. A critical mass represents a tipping point at which the rate of adoption dramatically increases and reaches self-sustaining growth. At this point, growth is associated with the network effects that adopters exert reciprocally. Although the point of critical mass is unique for each technology, it is suggested to occur between ten to twenty percent of adoption [29]-[30].

Blockchain's potential to transform society is often compared to the disruption brought about by the internet [6], [17], [7]. The internet and blockchain share common characteristics: both are open, distributed systems and both innovations unlock new economic value by lowering the cost of connections in the case of internet or of transactions in the case of blockchain. Also, blockchain is considered a significant component of the web 3.0 [6], [1]. Some go as far as saying that blockchain is the embedded economic layer the web never had [7].

Despite the ambitious expectations to disrupt the financial system, bitcoin's volatility has prompted discussions about its intrinsic value [31] and raised concerns over its ability to store value [32]. Also, the rate of failure of blockchain projects, at an astonishing 92 percent [12], combined with the 90 percent value drop in 2018 of some cryptocurrencies, reflected a period of disinvestment and disenchantment [33]. Blockchain technologies are expected to overcome this Trough of Disillusionment phase by 2021 [34].

Dynamic processes of adoption, such as the Bass model, can be used to simulate scenarios and assess the effectiveness of certain strategies [35]. For example, in early stages of adoption, as it is the case for blockchain, advertising has a strong impact. As the number of adopters increases, advertising appears to be overshadowed by word-of-mouth, arguably due to social influence or network effects that represent a growth engine in the rate of adoption, independent from the actions of the firms.

Innovation adoption research covers well the factors that influence the rate of adoption and diffusion [36]. Among the factors for IT adoption, top management support appears to be a good predictor from an individual and organisational perspective. Other factors are technical experience, perceived usefulness, behavioural intention, external pressure, organisational size and the expected return on investment [30], [37]. Naturally, these factors are to be used with caution, as they often build upon the assumption that individuals are rational when evaluating the usefulness of the technology.

Overall, it took three decades for the internet to move through the four phases of adoption and reshape the economy. Given the lessons learned from the internet and the pace at which information is being shared today, the tipping point for mass adoption might be reached by 2025 [10].

C. Platform Perspective

1. Strategy

The transition from pipeline to platforms business requires three strategic shifts: from resource control to resource orchestration; from internal optimisation to external interaction; from a focus on customer value to a focus on ecosystem value [38]. Blockchain-enabled platforms strategies also call for a sense of community to create and capture value outside organisational boundaries [39]. However, a blockchain-enabled platform would have one main difference with traditional platforms: there is no central orchestrator.

Network growth and changes in the environment call for periodic revisions, as the difficulty to coordinate an ecosystem rises and with it the risks of failure. In this sense, strategy formulation should be conceived as an iterative process rather than a linear process [40]. Managers should avoid defining too narrowly the roles of participants and embrace creative agility to pursue new ideas, reflect and adjust [41].

Blockchain-enabled marketplaces are characterised by greater competition and lower barriers to entry and innovation, allowing network participants to co-create a shared infrastructure [25]. In traditional models of innovation, internal R&D capabilities represented a competitive advantage and a key barrier to entry [42]. Inter-organisational collaboration can benefit from spreading the costs and risks associated with intensive R&D innovation projects. However, scholars suggest that firms have often missed the opportunity to use it as a source of competitive advantage. Companies have failed to achieve superior performance by focusing on outsourcing instead of building collaborative capabilities [43], [44].

Overall, blockchain-enabled platforms strategy is to be framed with a focus on networks rather than individual firms and should emphasise the co-creation of value [45]. In this sense, blockchain not only changes business models within corporations, it does so also at collaborative networks level [26].

2. Business Models

According to platform ecosystem strategies, building better business models is preferable than getting first to market [46]. Disruptive innovation theory also gives strategic importance to business models measuring the disruption phenomenon relative to the business model of another firm [27]. For example, Apple's mobile market domination is often associated with building a platform within a traditional pipeline business [38].

Disruptive innovation research provides useful recommendations for managers developing blockchain-enabled platforms [28]. On one side, disrupters are advised to focus on building business models highly differentiated from those of incumbents, rather than on building the right product. On the other side, incumbents facing disruption are encouraged to create new business lines designed to pursue the opportunities that arise from disruption. Last but not least, the incumbents' ability to manage and keep apart two different business units is argued to be critical for success [47].

The design and implementation of new business models while maintaining existing ones can be challenging for incumbents. As the disruptive theory advocates, the new offering can threaten the existing one and this can also occur for business models. Scholars have studied the benefits of ambidextrous organisations, characterised by higher degrees of agility and the ability to share a common vision and culture while maintaining separated explorative and exploitative units [48].

The ambidextrous approach [49] is conceived as a potential solution to the innovator's dilemma and a recipe for organisational adaptation and survival. In this context, senior managers appear again to be essential for a successful implementation, in particular: to articulate a common vision and a compelling strategy, to secure support from the management board and to manage inconsistencies or conflicts.

As suggested for the strategy, the appropriate business model should be flexible and scalable, following trial-error dynamics to accommodate for changes [50]. The lean startup approach, a term coined by Eric Ries [51], aims at optimising resources by using smaller and faster interactions for putting one's assumptions to the test with customers. The ability to integrate blockchain into existing workflows as well as the interoperability of platforms will prove critical to long-term scalability [52].

Among business models available in the literature, the business model Canvas allows for the development of alternative or radically new business models. In particular it can be applied to linear and non-linear business models including platforms, facilitating the transition from pipeline to platform businesses, which is vital for incumbents [2], [53]. The business model Canvas can be used as a strategic management tool helping to capture, visualise, communicate and enhance both strategic discussions and idea generation [54], [55].

Empirical studies have also revealed that users are combining it with other methodologies such as: SWOT analysis, blue ocean strategy, lean startup and balanced scorecard [55]. One reason might be that the business model Canvas fails to capture key strategic factors for business platforms such as core value alignment, governance, trust, matching mechanisms or network effects [2], [26].

3. Stakeholder Management

Blockchain platforms and applications entail the collaboration of multiple actors at different layers of the architecture. The size of the network is perceived as a determinant factor increasing coordination complexity [19]. Individuals perform better when they feel accountable for their actions and believe their voices matter, reducing the risk of free riders. The combination of capabilities is a compelling argument to overestimate the potential for value creation and therefore performance expectations, while challenges are often diminished and "can seem like someone else's problem" [40].

Drawing a parallel with the known minimum viable product (MVP) introduced by the lean startups approach, a minimum viable ecosystem (MVE) would be conceived as the "smallest configuration of partners that can be brought together for a healthy minimum viable network that still creates unique commercial value" [56], [57]. By removing traditional hierarchical controls, studies reveal that an increase in commitment and responsibility for individual actions is commonplace. However, potential downsides of group dynamic such as pressure to reach consensus, fear of judgement or topic fixation can be intensified as control is less apparent but more intense [58]. Since group bias increases the risk of groupthink [59], the diversity of actors contributes to enriching the blockchain ecosystem.

The sustainability of collaborative networks depends also on the alignment of core-values and goals among members throughout the different phases of the platform. The design of collaborative networks core-value maps can facilitate the following: the selection of potential partners with aligned values, the identification of incompatibilities among members to support conflict resolution and the impact assessment of the addition of new members to the network [60].

"Successful innovation requires tracking your partners and potential adopters as closely as you track your own development process" [40].

Last but not least, uncertainty and asymmetry of information among contributors on a blockchain-enabled platform can alter individuals' perceptions and expectations. Misalignments in the ecosystem can lead to unintended adjustments and therefore a deviation from the strategy. The high interdependence of actors also implies that speed-to-market

would be valuable only if partners are also ready, as bottlenecks may arise out of reach [40]. Managers are advised to assess the likelihood for parties to under-deliver, to understand the risk exposure and the underlying reasons -regulation, wrong expectations, incentives, lack of management support- and develop contingency plans.

4. Governance

Users interacting and transacting in traditional platforms, Uber or Airbnb for instance, rely on the platform owner's ability to set rules and processes to enable interactions in a safe environment. In blockchain-enabled platforms, trust relies on the underlying protocols like consensus mechanisms and the traceability of transactions. For the first time in history, users can reach consensus and coordination with no need for third-party intermediaries, allowing potentially for more granular and personalised services [8].

Trust and governance are therefore highly interlinked. Blockchain systems can be perceived as a form of governance, often referred to as decentralised autonomous organisations, a self-organised system as evoked with Coase's theory of the firm [23], 24]. These self-managed teams reach consensus about the platform's core values, design a shared mission and develop common practices. Governance and security deficiencies undermine users' perception of blockchain-enabled platforms' truthfulness and legitimacy [61]. Critical governance decisions have been around forks -"when someone takes the code of an existing application and uses it as the basis for a new application"- that created a competing structure and weakened the reinforcement effect resulting from the network effects of the original platform [62].

The governance should take into account the degree of complexity of the ecosystem [19] but also the degree of openness of the solution architecture [63], [8]. The latter would encourage both users and developers to contribute to the value creation. The main differences are as follows:

Permissionless networks are by definition uncensorable and therefore the governance could not provide sanctions, access or withdrawal rights. At the same time, as anyone can join the network and contribute, incentive mechanisms such as token rewards will have more importance to increase user engagement and avoid misbehaviour, excess or low quality of content that would destroy value.

Permissioned systems would benefit from appropriate levels of network coordination, control, compliance and human intervention. The governance of blockchain-enabled platforms in this case should contemplate the rules of functioning and provisions for dispute resolution, sanctions and access rights. Consortia formation offers advantages for managers as they are able to restrict the network to trusted nodes and members, however these less formalised relationships involve substantial trust-building, persuasion, shared vision and the alignment of expectations.

5. Network Effects

The consumer's utility associated to a product, physical or virtual, derives from its intrinsic value and the networked utility that results from the number of users in the network [64]. The literature recognises that network effects alter user behaviour and rationality [64], as the very nature of network systems reduces users' willingness to switch between platforms "unless everyone else does" [62]. This can also occur at the expense of adopting a superior technology. For example, the dominance of bitcoin reveals how competition from alternatives is not as effective due to the first mover advantage [30]. This reciprocal interdependence of users creates a reinforcement effect that was introduced in the adoption process.

The ability to bootstrap and operate a decentralised platform is referred to the cost of networking [25]. A native token is often used to crowdfund platform development and following that phase, incentive systems determine the conditions under which contributors are rewarded for providing resources such as computing power, applications or content. The cost of networking is therefore considered as an enabler for blockchain-enabled platforms to scale.

Although the customer base of incumbents' can be perceived as a deterrent, startups can benefit from a reduction in the cost of networking. Also, blockchain-enabled platforms can offer incentives for participants on the platform, by sharing the rewards resulting from direct and indirect network effects more fairly than in centralised platforms [25].

New blockchain-enabled platforms will have to face the chicken-and-egg problem that is inherent in all platform businesses but also characteristic of immature markets such as cryptocurrencies [62]. This chicken-and-egg scenario brings a new type of network effects where tokens have the potential to overcome the bootstrap problem, offering a greater financial utility for early adopters when the application utility is lower [65].

The literature suggests that managers willing to assess the value of a blockchain-enabled platform over time can rely on Metcalfe's Law, which states that the number of users in the network is measured by the number of unique addresses participating actively [66]. In this sense, Metcalfe's Law defends that the utility value of the network is proportional to the square of the number of users of the system.

Metcalfe's Law has served to diagnose bubbles and crashes in bitcoin [67]. These bubbles resulted from unsustainable, greater than exponential growth, leading to a correction of the value. Temporary value bubbles occur when the value of the network is not justified with a particular development that increases the value of the network, or an accompanied growth in the number of users participating in the network [66].

Because digital networks give rise to similar network effects as those traditionally associated with physical networks [68], one could expect a winner-takes-all dynamic [69], [70]. However, since participants' contribution on a blockchain-enabled platform is proportional to their stake in the platform, minorities that disagree with a suggested change face low lock-in and can be incentivised to fork at any time to create a separate platform [25]. The network split however questions the ability for platforms to scale, converge and ultimately experience a winner-takes-all dynamic.

Overall, managers are advised to focus first on the value of the interactions and assess the nature of network effects the platform is subject to [70]. Particularly in the context of blockchain-enabled platforms, managers need to find ways to ensure that the openness is translated into sources of value creation, rather than creating noise that would hinder user interaction [38].

D. Technological Perspective

1. Blockchain Architecture

Blockchain is conceptualised as a multi-layer communication and storage system [71], [72]. Important characteristics are: first, these layers are interdependent as participants acting and deciding at any of these layers are influenced by the decisions taken in the underlying layers; second, blockchain runs on top of internet, and is hence completely dependent upon the security of the latter.

There are multiple actors involved on a blockchain-enabled platform: architects to design the platform, developers to create applications and smart contracts, network operators, blockchain regulators and users. The challenge stems from the fact that, unlike in other IT architectures, actors in the system do not necessarily know or trust each other, yet they all contribute to the network in exchange for value [73].

Different blockchains result from the different approaches to access rights, consensus mechanisms to record creation and validation, and incentive mechanisms [22], [72], [74]. Despite a lack of consensus in the literature, blockchains tend to differ based on choices made across two dimensions:

- First dimension: access
  - **Permissionless** blockchains, such as Bitcoin or Ethereum, allow anyone to join the network, create a record and act as a verifier. However, these blockchains suffer from low throughput, scaling limitations and high energy consumption.
  - On the other hand, **permissioned** blockchains require the authorisation from designated parties to be able to create a record and transact; whether a consortium, such as R3 Corda, or a unique entity, this is a compromise to the decentralization promise of blockchain.
- Second dimension: transparency
  - **Public** blockchains allow anybody to access the information and participate pseudonymously.
  - **Private** blockchains allow data to be shared only with trusted parties and offer better privacy.

Whilst permissionless blockchains are often associated with public, and permissioned associated with private [26], there are use-cases for all four combinations. Examples of public permissioned blockchains can be found in supply chains, or in annual corporate reports where only legitimate participants can add information, but everyone can access it.

Further to that, the technology imposes some trade-offs known as the scalability trilemma where blockchain systems can have at the most two out of the three following properties simultaneously: scalability, security and decentralisation [75]. Blockchain developers can navigate among these trade-offs to personalise the blockchain solution, leading to multiple possible combinations of hybrid blockchains [76]. These hybrid systems have emerged to respond to often conflicting requirements and balance the existing trade-offs [72], [77]. For example, applications in the financial services demand a high level of transparency for reporting and regulatory reasons whilst requiring strong guarantees on the privacy of their customers' data.

2. Blockchain Applications

Blockchain in particular and DLTs in general have received great attention from the "tech" community, industry practitioners, scholars, policymakers and the media. Although over promotion can be counterproductive for the long-term development of a new technology [78], the number of blockchain applications is growing at pace [79]. There are many potential use-cases beyond traditional financial transactions such as digital identity, voting systems, supply chain, personal data monetisation and remittance services [80], [20].

Although it remains difficult to predict what use-cases will have a lasting impact, and if and where they will converge, the development of blockchain since 2008 can be split in four phases [7], [81]:

- **Phase 1**: focus on economic efficiency and costs savings for decentralised transactions of currencies. This phase saw an explosion in the number of cryptocurrencies created and traded.
- **Phase 2**: 2015 saw the release of Ethereum, a turing-complete distributed public blockchain network that offered for the first time the ability to create fully functional smart contracts. Smart contracts are computer code that can technically enforce a contract between multiple parties, without the need to trust the parties nor a trust broker [82]. The Hyperledger foundation followed to offer blockchain solutions for enterprises with a series of frameworks and tools. This second stage explores decentralisation beyond financial markets and the transfer of digital assets beyond cryptocurrencies.
- **Phase 3**: focus on the development of decentralised applications, and in parallel, research on how blockchain can reduce organisational boundaries and provide fairer transnational governance structures to generate greater value.
- **Phase 4**: the intersection of blockchain and other technologies such as artificial intelligence (AI) and internet of things (IoT).

Paradoxically, the fact that the community is structured around the idea of decentralization, makes it incredibly difficult to pursue, and hence reach, any form of consensus on taxonomies, let alone standards. This is probably the main obstacle to the democratisation of the technology at this stage, as it was for Internet in the 80s [83].

Overall, the literature reveals that the diversity of applications for blockchain explains the diversity of opinions about the value and potential of the technology. In this sense, no one solution fits all. The theories of adoption and diffusion suggest that adoption occurs in phases that can be fostered or hindered by multiple factors. The pace of evolution and associated uncertainty around blockchain require a continuous adaptation of the strategy of blockchain-enabled platforms to changes in the environment. One important takeaway is that companies building collaborative capabilities can and should use these as a source of competitive advantage. Also, platform managers willing to embrace the technology and rely on existing business models, can turn to the business model Canvas. Last but not least, the literature review revealed the lack of a framework to guide those building blockchain-enabled platforms.

E. Research Focus

The hype around blockchain, stemming mostly from its promises of global disruption, has resulted in more than half a million new publications between 2016 and 2018 [74], from white papers to blog articles and semi-scientific contributions. Scholars have only begun to look at blockchain and so far, mainly focussed on three themes: demystifying the technology, its disruptive potential and identifying specific use-cases [19].

However there remains many open questions for incumbents and startups willing to embrace blockchain technology. The high rate of failure of blockchain projects is indicative of a real struggle to leverage the technology for commercial success. There is a need for research on the main challenges blockchain-enabled platforms are facing today and how these challenges differ for incumbents and startups. This is the focus of this paper.

III. Methodology

A. Research Strategy

The empirical research in this paper focuses on the challenges that blockchain-enabled platforms currently face. While a case study strategy would offer valuable insights, it would fail to provide a holistic view of the challenges faced by stakeholders, given their diversity and given the divergent nature of the technology.

A survey strategy [84] was employed for the purpose of this research because it yields extensive and detailed information to interpret and draw representative results about the population. Given the limited empirical work available, the author of this paper followed an inductive approach with a multi-level design to reveal the challenges at the platform-level as well as the dependencies at inter-players levels [85].

B. Data Sources

The sources for data collection used are threefold: first, in-depth semi-structured interviews; second, emails, observations and follow-

up interviews; third, internet resources. The triangulation of data from various sources strengthened the robustness of the findings [86], [87].

Acknowledging the limitations of closed-ended questions, characteristic of questionnaires, the author opted for in-depth semi-structured interviews [88]. This type of qualitative research allowed to unveil the root causes of the research question and provided for flexibility to formulate further questions that would explore specific themes based on the experience of the informant [84], [85].

C. Data Collection

Considering that the researcher had no personal ties with the informants, studies confirmed the effectiveness of introductory messages and field conferences [89]. In this sense, the author of this paper approached leading actors through LinkedIn explaining the objective of the research and why the experience of the informant was relevant. Furthermore, blockchain and fintech conferences proved to be the ideal setting to network and collect data through observations where leading actors discussed the potential and challenges of the technology.

Following a multi-stakeholder approach, twenty semi-structured field interviews were conducted with leaders in blockchain and platform strategy, taking into account ethical and privacy considerations [88]. A summary of the informants' profile, gender and geography can be found in appendix A. The diversity of the group aimed at providing a comprehensive overview. All interviews were conducted through web conferencing, video or call, an environment conducive to productive discussions without interruptions. On average, interviews were an hour long, ranging from forty to ninety minutes.

D. Data Analysis

The twenty interviews were recorded and transcribed. As suggested for inductive analysis, the information collected was organised, disaggregated into categories making use of mind maps and tables in order to highlight patterns and differences among the different interviewees [90]. Also, the comparison with the existing literature allowed to uncover similarities and differences, demonstrating the need for further research.

IV. Findings

This section sets out the findings from the twenty interviews conducted with leading blockchain-enabled platform actors. As with the literature review, the findings have been analysed from three standpoints: innovation, platforms and technology. The interviews allowed for the identification of nine main challenges that blockchain-enabled platforms are currently facing.

A. Innovation Perspective

- **Challenge 1**: the innovation potential of blockchain is debated.

In innovating with blockchain, incumbents and startups mostly pursue different objectives, respectively efficiency and disruption. Consultants and entrepreneurs agree that large organisations face difficulties to understand what blockchain is and think about it in innovative terms. There are multiple reasons for that. First, the centralised nature of the business makes it difficult for senior managers and executives to think decentralised. Second, large organisations tend to prioritize improvements in efficiency over risky innovative paths: in most incumbents, under-resourced IT departments have little time to explore new ways to do their job, let alone to do business. A blockchain service provider explains: "*What limits large organisations is that they are stuck in the traditional way of doing business*".

As a result, most incumbents deny the disruptive potential of the technology and recognise that their investment, if any, is mainly for efficiency purposes. Paradoxically, the very same executives argued that embracing new technologies such as blockchain is a necessary step for survival. Furthermore, incumbents tended to praise their own innovation efforts, legitimizing their investments with long-term positive spillovers. A financial executive said of the innovation efforts of his company: "*We are innovating like crazy and the other great news is that because we already have the network, we can roll it out and make it backward compatible. I am really looking at cross-border payments in our banking community for another 100 years of resilient and good experience for customers*".

Startups, on the other side, tended to focus on finding new ways to create and capture value; much enthusiasm was shared, for instance, around the potential for blockchain to unlock the data economy, the access economy and ultimately the circular economy. An entrepreneur explains: "*I think we are moving into a new world war, and while we all know it will be fought on the internet, few realise the reason for the war will be the internet of value". Blockchain, they argue, could for instance provide users ownership over their data and eventually remuneration for its use*.

The fact that incumbents and startups fail to agree on the disruptive nature of blockchain is a challenge for it limits cooperation, fosters fragmentation and ultimately hinders innovation. The risk for entrepreneurs is to fail to gain sufficient market share, whilst the risk for incumbents is to miss the train and find themselves two or three innovation waves behind, at which point they will no longer be able to catch up.

- **Challenge 2:** the mass adoption of blockchain-enabled platforms is constrained by endogenous challenges such as crypto volatility and digital divide.

Experts agree that the cryptocurrency market is perceived as an indicator of the health of the whole ecosystem that leverages blockchain, and the volatility of cryptocurrencies negatively affects the adoption of blockchain-enabled platforms. The hype around cryptocurrencies in 2016-2017 attracted speculators, who further increased price volatility, creating a speculative bubble. The 2018 crypto market downturn balanced the type of participants in the ecosystem. The ability for investors, users, enterprises and governments to use and accept cryptocurrencies as a method of payment is seriously undermined by the high volatility of prices. Furthermore, mass adoption is constrained by the fact that cryptocurrencies still cannot really be used as a mainstream method of payment. This situation creates a vicious circle where a low market acceptance of cryptocurrencies is translated into low market liquidity. An entrepreneur explains how tokenization can have a positive spillover increasing liquidity within the ecosystem: "*For our project to scale and be successful we required liquidity, the one benefit from doing it with tokens is that it becomes accessible to a lot of people, and people can exchange and can keep it, hold the token or not*".

From a user perspective, price instability impacts buyers' purchasing power and behaviour. From an enterprise perspective, volatility hampers businesses' ability to predict revenues and invest accordingly. Even blockchain entrepreneurs who raised capital through Initial Coin Offerings (ICOs), saw much of their capital "burnt" by the market itself. Volatility damages investors' confidence and hinders long-term decentralised applications that require price stability. As an entrepreneur warns, "*Cryptos are considered the most dangerous investments and it is not a surprise that regulators are trying to protect investors. Hopefully cryptos will stabilise, otherwise there will be huge problems for mass adoption*".

While blockchain-enabled platforms are becoming more common, they are still for "tech-savvies". Experts recognise that despite the increase in adoption since 2016, the diversity of users in blockchain-

enabled platforms remains the same: users are individuals that are fairly familiar with cryptocurrencies and have a particular interest in decentralisation. As an entrepreneur describes, "*we haven't really seen normal users, almost everyone that uses the platform, understands what Bitcoin is*".

From an end-user perspective, experts argue that crypto-illiteracy would be a barrier for adoption, and so advise that the emphasis should be placed on the advantages the technology offers such as traceability, transparency and immutability. An entrepreneur explains: "*What we want is to sell our solution without talking about blockchain, our customers mostly don't care about using blockchain or not. We don't want to scare people by using some scary words, such as cryptocurrencies or blockchain*".

From a business perspective, experts highlighted the difficulty to understand blockchain and to develop sound use-cases to what consultants and Blockchain-as-a-Service (BaaS) appear to be essential. These organisations have multidisciplinary teams that provide support for startups and large organisations at different levels: requirements drafting, strategy definition, blockchain development, and last, training and maintenance.

- **Challenge 3**: the mass adoption of blockchain-enabled platforms is constrained by exogenous challenges such as regulatory divergence and access to vital financial services.

Blockchain-enabled platforms are constrained by divergent and sometimes conflicting regulatory approaches. A barrier for platforms in terms of development and liability is that in some regions, platforms are not able to operate due to regulatory uncertainty. A risk arises for platforms when users circumvent national regulations in place. In response to the legal vacuum, crypto-friendly countries such as Switzerland, Malta and Liechtenstein, have tried to attract crypto and blockchain businesses. These countries have adapted state policies to increase transparency and regulate the activities in the blockchain and crypto markets. However, the practical implementation of regulation adds more complexity for blockchain-enabled platforms as it is bound by national jurisdictions while blockchain is a global phenomenon. A legal expert explains: *"The ICO guidelines provide a small framework but then [regulators] said that the existing laws can be applied to blockchain. The real dilemma is how to synchronise the existing legal framework with a brand new technology and bring both together to build something"*.

Another key external difficulty for startups building blockchain-enabled platforms is access to critical financial services. Some blockchain entrepreneurs have agonised as banks have repeatedly rejected to support their projects. Part of the reason banks are giving crypto startups a hard time is the difficulty to comply with heavy regulation, from tracking the origin of the funds raised, to taxing profits, or lending given the high-risk profile of crypto markets. An entrepreneur explains: *"Banks don't want to get involved for fear that there is money laundering going on in crypto, the amount of work they would have to do to ensure a client's money is clean outweighs what they would earn in fees"*.

Businesses, particularly in the financial sector, have fiduciary duties and responsibilities. Complying with "Know Your Customer" (KYC) and "Anti-Money Laundering" (AML) procedures is a challenge for businesses as these are not embedded in the blockchain. With that being said, a bank executive explains how the dilemma is not with the technology: *"It is not a blockchain problem, the issue is that the projects themselves need to comply with the regulation. This can be solved by making whitelists of investors that comply with KYC/AML procedures. There are companies that are doing this very well"*.

Even entrepreneurs who managed to raise capital without using traditional markets, through ICOs, faced difficulties. Those that did survive the crash of the crypto market interestingly face legitimacy issues, as explained by an entrepreneur: *"ICO booming in 2017 brought a lot of attention, notably bad attention, as it is now estimated that at least 80% of ICOs in 2017 were scams and failed to deliver on their promises. Since we have raised this huge amount of money at the very specific time where there were so many scams, so many bad projects, we kind of had to prove that we were not a scam ourselves"*.

B. Platform Perspective

- **Challenge 4:** misguided and misguiding managers lead many platforms to fail.

This challenge is both a platform and a technological one. For most entrepreneurs interviewed, the decision to develop a blockchain platform has been influenced primarily by curiosity and experimentation, less so by a clear view on the returns. Larger businesses also struggle with a similar challenge, shareholders or the executive team asking senior managers to develop a platform based on a technology they barely understand. A consultant explains: *"Senior executives did not have the time to learn or understand what blockchain is and is not, often pressed by the engineering departments who would put more emphasis on the technology than the returns for the company"*.

Platforms and blockchain are considered still buzzwords. As a result, both remain misunderstood by most people, hampering their ability to properly evaluate the technology requirements for specific platform use-cases. The high level of interest for platforms and blockchain technology lure enthusiasts and enterprises into the trap of defining a solution before the problem. As a blockchain consultant puts it: *"the biggest challenge for enterprises is to find their way into blockchain and into platforms, because they don't know what kind of problems these can solve today. For still too many people, the sole purpose of building a blockchain platform is to tell the world you are using blockchain"*.

Choosing the right blockchain technology stack for the right platform requires a thorough case-by-case assessment of the requirements. The fact that end-users and businesses often have opposite requirements when it comes to blockchain-enabled platforms, is a serious challenge for developers if they are not clear on what problem they want to solve. Experts interviewed recognised they had too often considered the requirements from the wrong perspective.

Not surprisingly, consultants confessed that most of the time, proof-of-concepts revealed that blockchain was actually not the optimal solution and that other technologies could solve the problem at hand better, faster, and for a fraction of the investment.

The hype around blockchain leads many to try to solve the big problems before the little ones. Finding the appropriate size and scope of a platform to solve a particular problem can be challenging. Bitcoin has proven that blockchain is a global network that can scale up from a business functionality point of view, yet it has also proved that it is far from being able to compete today with existing payments systems. In that sense, experts agreed that at this early stage, a global and large platform is a mistake and recommend to start small.

Another consultant emphasises the importance of speed-to-market: *"If you have a problem, validate it, build something with it and release it in less than three months and if you can't do that it means that you have to narrow down the view to something very specific until you are confident you can build it in three months"*. This is linked to challenge 8 hereafter: as time-to-market is shrunk, security becomes a second priority whilst it is paramount to trust, and trust is itself paramount to blockchain users.

- **Challenge 5**: blockchain-enabled platforms face new and unique governance issues.

Conflicts of interest can quickly arise in blockchain-enabled

ecosystems. A profit-oriented platform, leveraging on blockchain or not, often has a set of obligations towards its investors such as dividends and governing or voting rights. Although during the early stages of adoption, the interests of the users are generally aligned with those of the platform, as the network grows and captures more value, the need for maximising the financial returns for shareholders prevails. This scenario does not necessarily translate for blockchain-enabled platforms. A blockchain entrepreneur of a fully decentralised blockchain-enabled marketplace explains why he wants to separate the foundation in charge of the platform from the company that initiated the development of the project: *"We want to split the foundation and the company because the company is very incentivised to not decentralise whereas the foundation's goal is to make sure that the protocol succeeds. We are trying to separate those concerns; we have a lot of those discussions internally if we want to make money or not. And that shouldn't be a decision we make. So, I think that we just want to get away from that. From a company perspective, it doesn't make much sense, but we want to succeed whether the company does or not"*.

Interestingly, another entrepreneur explains how he managed to keep his community together by leveraging a utility token sale that allowed investors to participate in the network and at the same time redistribute value equitably among the different actors: *"those that invested during the ICO, we call them "our" community and not shareholders as we don't owe them anything, legally speaking. They have bought the product we are working on to be released. We have a very transparent way of communicating our advancements and what the money is used for"*.

For incumbents, the shift from "ego-systems" to ecosystems will take time. Large organisations are having difficulties moving from a silo perspective to a multi-stakeholder perspective as they often have their own expectations of what the solution should look like. Moreover, participants in a blockchain network are often required to wear many hats as they can have different roles in different platforms and a separate mandate within their own organisation. A key challenge is how different parties or even competitors can work together to build an ecosystem and align their objectives and core values for a specific purpose.

Within a blockchain ecosystem, collaboration aims at creating and enriching the network while these parties might compete at the application level. Although a founder would set up the first node to start a network, in a distributed model there is no central ownership of the platform. In this sense, participants on the network need to collaborate around the design, governance and monitoring of the platform. The community model of management relies on the wisdom of participants on the platform to make the right decisions to make the network secure, stable, resilient and accessible. A platform expert explains: *"we call it Platform 2.0, the combination of blockchain and AI to do cognition at scale, or so-called collective intelligence, to be able to jointly address global challenges"*. For the time being though, for blockchain-enabled platforms to scale, governance will need to rely on more than participants' wisdom or AI.

Directly linked to the governance model is the size and composition of the ecosystem, and there is little guidance available on the topic. The challenge is to decide with who to build the platform and whether it is pertinent to maintain some kind of privileges or superiority for the work already conducted, as it may disincentive other partners to join the network and undermine the benefits of blockchain.

All these difficulties converge into one of the most difficult challenges for blockchain-enabled platforms: decentralized governance. Governance, in the case of blockchain, encompasses the definition of rules and norms that determine the access rights, transaction validity, conflict resolution, issuance of new assets and tokenization. The agreed governance is implemented on-chain, hardcoded for instance in consensus algorithms that are aimed to be self-governed. Nevertheless, governance also occurs off-chain. Even permissionless blockchains such as Ethereum rely on an off-chain governance process for a variety of decisions such as proposals for additional features. Examples of off-chain channels are the Ethereum Github webpage or the Ethereum Community Forum. On-chain processes require the validation from stakeholders; off-chain processes may not. Managing these processes in parallel, in a fully decentralized manner, is proving to be a major obstacle to the growth of blockchain-enabled platforms.

Blockchain experts mostly agree that defining governance models should precede technological choices as the latter implement the rules and internal processes agreed upon. Surprisingly though, multiple experts interviewed confessed they had not yet a clear view on governance, rules and conflict resolution. In two cases, governance was delegated to a community of volunteers. An entrepreneur that launched a platform two years ago argued: *"The things that we are focusing on right now are outside of the governance process. Governance process in decentralised apps is quite new for us and we haven't decided yet how it will be structured. Right now, we are putting together a foundation board, people that we know are coming from different areas and that would serve as a fair and diverse community"*. Indeed, as the off-chain governance process can prove to be challenging, particularly in public networks, it is often delegated to foundations or communities of volunteers.

"It's been 4 years and we are still trying to solve the same problems that we thought of when we started. It is really hard, not just the technology but the way people act it is different from centralised networks and it is very hard to police them. So getting people to do the right thing is hard", concludes an entrepreneur.

- **Challenge 6:** incentives strategies and mechanisms for the launch, development and operation of blockchain-enabled platforms are still exploratory.

The economic framework of any ecosystem takes into account the different motives of participants. As highlighted before, these may differ within and across blockchain-enabled ecosystems. In permissionless networks in particular, economic incentives are fundamental as these networks rely greatly on the contribution of users to the value generated within the network. Tokenization is one of the ways in which blockchain is disrupting traditional businesses: crypto-tokens are a digital form of value that allow incentives to be kept aligned between network participants and foster network growth. Although tokens were previously used mostly to raise funds in ICOs, most tokenization is expected in supply chain, loyalty programs and membership rights.

The challenge is how to incentivise participants and ensure that the creation of value is maximised and the destruction of value minimised. There is not one model fits all, notably given that different platforms have different stakeholders, like curators and moderators. Their role is essential for blockchain-enabled marketplaces to compete with giants like Amazon on the relevance of their listings per customer. An entrepreneur shares concerns about the potential spillover effects of growth in self-governed organisations: *"Right now it is very basic, you just take whoever but I think that if it gets to millions of users it would be hard to decide who you wanna use as a moderator. Moderators that are more responsive, more trustworthy, with better reputation may become expensive"*.

An important component to incentivising stakeholders to join one's platform is communication. The hype around blockchain technology led to the publication of hundreds of ambitious blockchain-platform white papers, well supported by marketing plans and little more. Whilst these projects did capture enough attention to raise capital through ICOs, the majority failed to deliver tangible results let alone returns for

initial backers. Experts criticised that many startups often rush out the process of ideation, formation and implementation when they fear they run out of cash. An entrepreneur explains: *"One of the first mistakes that startups make when they get funded is to over-recruit, purchase infrastructure, travel everywhere and all the sudden there is nothing left to develop the product"*.

One last challenge concerning incentivising users to join a platform is community support. Projects with greater decentralisation characteristics resonate better with the intrinsic motivators of blockchain evangelists and other "crypto-lovers", who in turn voluntarily support the projects on social media, in conferences and in community forums. Despite the lack of a well-defined strategy, these platforms appear to grow organically, relying on the power of word-of-mouth. The founder of a public permissionless blockchain-enabled platform confessed: *"We never did any formal marketing or advertising efforts until the last few months other than just us tweeting. The way that we built it up was to focus on getting developers to build the platform but kept on engaging with people within the crypto community to continue inviting people to join the project. I think that because the idea was kind of crazy, we got a lot of trust early on which brought a lot of attention to what we were doing"*.

The challenge is that the opposite is also true, and these actors often badmouth projects they disagree with, ultimately undermining their ability to raise capital or develop their customer base.

C. Technological Perspective

- **Challenge 7**: the skills required to develop blockchain-enabled platforms are diverse, scarce, and under stress.

Blockchain is still in its infancy. This implies first a shortage of blockchain developers in the world. Although today, large organisations often have their own innovation labs, the skills shortage makes it close to impossible to find, never mind hire, experienced blockchain developers. It is also worth noting that this is further exacerbated by the fact that the financial sector, one of the first in attracting these skills, has the means to retain talent. An entrepreneur explains: *"Finding people to write and test the code can be very expensive. They are expensive because they are only a few. Many experts who used to work for leading blockchain companies have been poached by banks"*.

Interestingly though, this lack of competence is fostering collaboration between BaaS companies, consultant services, startups and established organisations. The Enterprise Ethereum Alliance or the PwC and Microsoft alliances are notable examples.

Startups recognise that developers' motivators are not only extrinsic, notably financial, but also very much intrinsic. Blockchain remains one of these fields that attracts people pursuing other objectives, like technological breakthroughs, social or political impact. Attracting talent in such a competitive market is then even harder, as recruits, often millennials, look for cultural fit and social impact, not only for an attractive benefits package [91]. Last, given the volatility of the market and the disruptive potential of many projects, it is particularly difficult to retain talent as stock options may suddenly drop in value and competitors develop more appealing projects. The fact that blockchain is still in its infancy also implies that the technology is new, and changes often. As a result, blockchain developers spend time solving technical bugs without much experience doing so. A developer describes how challenging it is to stay up-to-date: *"I hope that we will not see more divergence because it is hard to find the proper technology in this kind of context. When we are developing a blockchain application we know that maybe in one or two years we will have to recode everything using the latest version because the way we coded will be obsolete. We are aware of that and we are taking the risk"*.

- **Challenge 8**: the security implications of blockchain are largely overlooked.

Experts tend to agree that successful blockchain-enabled platform development is bound to follow an iterative process: first, prototypes should be built as quickly as possible using minimum resources; second, potential customers should test prototypes and their feedback be used to improve the solution; third, the process should be repeated over and over again.

This poses a critical challenge, all the more important because it was not mentioned by any of the interviewees. The security of blockchain was never discussed yet it is of tremendous importance for ledgers are immutable, smart contract code is law, and the transparency of public blockchains makes its data accessible to all, including hackers. Security-by-design, or even better, security-by-default, is a growing concern and hence discipline in other emerging technologies such as 5G, IoT and AI. Indeed, many stakeholders worry that future mobile networks can be spied upon, that IoT devices may be hijacked or that AI will turn against humans... Yet all these technologies can be updated when a vulnerability is discovered. Not blockchain, at least not without a hard fork, which impacts trust considerably.

Unfortunately, too few worry about the security implications of sensitive data being stored on the blockchain, misuse of legitimate protocols to create money out of thin air, etc. As a result, blockchain security vulnerabilities have already led to significant loss of value, and bad press overall for the entire community.

Last but not least, the combined skills shortage in cybersecurity and blockchain is a serious growth barrier for the technology [92].

- **Challenge 9**: blockchain technology may not be ready yet to deliver fully on its key promise: decentralization.

For blockchain to become a unifying system to record all transactions, there are still many obstacles: today, information is managed in silos within organisations and often duplicated over and over again. The reconciliation of information across systems is time-consuming and error-prone. Blockchain is an enabling technology and the quality of the information it will provide will be as good as the data source. Thereupon, blockchain-enabled platforms' ability to scale up to full adoption is dependent on the ability for platforms to interoperate between themselves and with existing information systems. A digital transformation executive explains: *"blockchain needs to operate over homogeneous datasets. The challenge is that information systems have evolved independently with different standards. Even information systems using the same language are not managed homogeneously. Blockchain requires that these different datasets are coordinated so that they can connect to each other and understand each other. The challenge is how to turn heterogeneous databases into homogenous ones and make them compatible to exchange value"*.

Today, the return on investment (ROI) to migrate existing systems and processes to blockchain-enabled platforms fails to convince incumbents. Even after running several prototypes, for most, the perceived ROI was uncertain and rather long-term, notably because it is directly dependent upon the ability or commitment of other participants in the network to invest as well, so information can be exchanged. Experts interviewed argued that none of the available options today scaled enough to convince investors and the broader ecosystem. As an incumbent explained: *"Based on our experience so far, banks are not ready to implement blockchain as they would have to migrate and be interoperating with that new platform. For them it is much easier to implement it with APIs and our existing offer than it is with blockchain"*.

For decentralization to become mainstream, blockchain-enabled platforms may require centralised implementations first.

Incumbents focus primarily on performance, privacy, scalability and the ability to transact only with interested parties. In that sense, the

additional complexity and costs associated with creating "artificial" permissions on top of a permissionless platform outweigh the benefits. As a result, most incumbents are developing solutions on private and permissioned blockchains and often using IBM's Hyperledger or Corda. An incumbent executive explains: *"Hyperledger of course. We use smart contracts to execute payments' instructions. The buyer, the seller and the involved banks are able to transact ensuring that only the parties involved in the transaction have access to the details. You can't do that easily with Ethereum"*.

Naturally, on the other side, many startups hope to leverage decentralization to grow and challenge incumbents in no time. Their B2C applications try to benefit from the higher degree of security, anonymity, transparency and decentralisation of public and permissionless blockchains. And Hyperledger is obviously not the solution of choice, as explained by a consultant: *"Hyperledger, to my perspective, is too proprietary even though it is open source. If you think about the clients that do implement Hyperledger today, these are all premium IBM passengers"*.

There is a new generation of platforms bringing together private and permissioned with public and permissionless platforms. Kaleido, for instance, is a full-stack enterprise platform developed by ConsenSys that records only part of the information on the main chain of the public blockchain. In this sense, Kaleido offers developers the flexibility to select the most appropriate protocol to build on top of Ethereum and uses AWS infrastructure to scale up. The sidechain allows for greater levels of privacy and speed to test a new application. However, here again, there are contradictory opinions as an entrepreneur explains: *"for enterprises, blockchain is generally just a technology enabler while for some end-users, it is almost a religion. Balancing expectations is difficult: running decentralized blockchain applications on the most proprietary and biggest cloud on the planet can prove challenging from an end-user perspective even if it can make perfect sense from a business perspective"*.

Overall, the fast evolution of blockchain technology calls for agility. The ability to stay at the forefront of technological developments will depend as much on the collaborative capabilities among stakeholders as on the ability to go to market with fast-built prototypes and improve the solution on a continuous basis. Currently, it appears that centralisation may be a necessary step to take for the technology to converge and gain adopters. Migration and interoperability costs are currently perceived as major obstacles and regulation and standardisation have not really kicked in yet. If and when these barriers are lifted, the technology will face its ultimate challenge: shifting fully towards decentralization.

V. Discussion

The literature review and the diversity of profiles interviewed provided a qualitative basis for a holistic analysis of the key challenges inherent to blockchain-enabled platforms. This analysis, in turn, provides the basis for the following discussion, structured around two sections: first, a comparative analysis of the findings and the literature review; second, a set of recommendations for managers developing blockchain-enabled platforms. Also, a graphical abstract a theoretical framework capturing the key takeaways of this paper can be found in appendix B.

A. Comparative Analysis

Traditional centralised authorities are seeing their legitimacy increasingly questioned as blockchain technology matures. Libertarian blockchain evangelists praise the trustless and apolitical nature of blockchain technology, pushing to replace hierarchal centralised systems, reduce market power, privacy risks, censorship risks and give back ownership of data to their owners [62], [25], [16]. Yet as the findings unveiled, there are multiple and interrelated challenges that hamper this vision.

The findings confirmed what the literature suggested: there are divergent opinions on the disruptive potential of the technology. Some experts interviewed look at blockchain to improve existing processes, while others pursue much more ambitious goals, such as the internet of value. The disruptive innovation theory in the literature review also showed that startups can supplant incumbents by targeting new markets that remained ignored by the latter: given that blockchain is still in its infancy, no finding confirms this theory yet.

Findings also support the fact that adoption drivers differ between startups and incumbents. While the former are often looking into new business models, the latter are mainly embracing the technology for efficiency purposes, without questioning existing business models. In particular, incumbents often casted doubts on the ROI, given migration costs and uncertain returns. That being said, the findings also revealed that some large organisations, particularly in the financial sector, are actually looking at blockchain in great depth and have gone as far as insourcing their blockchain development efforts entirely.

Despite transparency and community support being prominent aspects in the blockchain space, the literature on decentralised governance is limited and part of the reason is the difficulty to capture and share lessons learnt [93], [94]. Experts also recognise that there are no successful examples with sufficient track record as of now. The findings revealed that for important decisions such as the development of additional features, governance in public blockchain-enabled platforms often occurs off-chain [95]. As these transactions do not occur on the main blockchain, the latter loses some of its core values points such as transparency. It remains to be seen how data exchanged off-chain will be managed and taken advantage of.

The literature reflected well the need for a lean startup approach [51] which emphasises the need for a MVP. However, the fast pace of change and the inherent trade-offs that the technology imposes, calls for temporary concessions, the findings revealed. It appeared that managers are accepting the introduction of centralised components for the sake of adoption, even on the essence of blockchain: decentralization.

Blockchain was designed to replace imposed confidence, or vertical trust, by voluntary confidence. Whether this happens or not, a transition period is needed to replace the existing trust models, notably on companies and processes. The findings revealed how many of these companies shy away for a variety of legitimate and questionable reasons. For instance, if the current volatility of cryptocurrencies is indeed an obstacle, price fluctuations will eventually stabilise as they did for tulips or gold [96], when more people get involved and liquidity increases. As a result, this argument is short-term at best, and does not affect the long-term potential of blockchain. Not recognising this potential is as much a risk as it is to invest in crypto-markets today.

The literature review also suggested that regulated institutions and established organisations would try to circumvent blockchain businesses to protect the status quo [6]. The findings indeed revealed that they do play a role in dismissing blockchain, discrediting the legitimacy of blockchain businesses. Although the first adopters were often associated with illegitimate activities, the Bitcoin economy has grown in size and scope with legitimate applications [97].

Also, the findings illustrate challenges that were not captured by the literature review. First, blockchain-enabled platforms often need to navigate uncharted regulatory waters for the service they pretend to offer, but also go to great lengths to secure basic but vital financial services for their very own corporate operations, from raising capital to paying salaries [98]. On the other side, the findings revealed that many blockchain projects with ambitious marketing plans were unable to control their expenses and quickly ran out of money.

Among the factors of adoption that the literature review highlighted, the findings revealed that external pressure and management support do not always lead to positive outcomes as managers of blockchain-enabled platforms also often rushed into development without a proper use-case assessment.

In this light and despite what libertarians may think, it seems that mass blockchain adoption is going to require top-down initiatives, even if only for a few years, to complement bottom-up work. This does not mean that bottom-initiatives play no role, to the contrary. Among adoption strategies, the literature review suggested that advertising would be more efficient during the early stages of adoption and word-of-mouth would gain more traction later in the adoption curve. The findings demonstrate that the opposite actually occurs in the case of permissionless blockchain-enabled platforms. As informants argued, permissionless networks grow more organically thanks to word-of-mouth communication, a good example of an influential bottom-up driving force. Indeed, participants in decentralised networks actively support blockchain solutions that are aligned with their core values.

The findings revealed how the technology is largely misunderstood, leading many to fail even to find a realistic use-case. While the literature and industry reports focus greatly on the characteristics of the technology and its potential application, experts emphasised that the development of a blockchain application calls for a careful assessment of the actual need of blockchain. In that sense, the focus should be placed on the desired outcome and value proposition, rather than on the underlying technology. In some way, the less blockchain is talked about, i.e. the more emphasis is put on the ground-breaking solutions it offers to traditional problems, the more the technology will become mainstream.

The literature review highlighted the absence of taxonomies, let alone standards. The findings recognised that a significant challenge for the community is to navigate ambiguity and diverging opinions, exemplified in the astonishing number of developments, projects and cryptocurrencies. For blockchain-enabled platforms to scale, reach a critical mass and reap network effects, a higher degree of convergence and standardisation appears to be a difficult, but necessary route.

Emerging technologies such as blockchain are characterised by a hype-curve associated with inflated expectations. The findings suggest that the period of disillusionment has helped reduce the number of speculative investors, leading to a "healthier" environment for the development of blockchain projects. The steady growth of blockchain applications, crypto-wallet owners and job openings [79], [99], [100] indicate that blockchain and the token economy are here to stay [78].

In short, the findings complemented the literature review substantively, shedding light on the challenges that blockchain-enabled platforms are facing today.

B. Recommendations

This section aims at providing practical recommendations with a dual focus on incumbents and startups based on the main development phases of blockchain-enabled platforms. The key components are summarised in Fig. 2.

1. Research Phase

- **Value proposition**

The very first consideration for managers thinking about developing a blockchain-enabled platform is to clearly define the problem to solve. Too often the hype around the blockchain and platforms lead executives and developers to rush into a solution. Blockchain and platforms are enablers, not ends in themselves. Blockchain developers and consultants should accompany their clients on refining the value proposition and empower managers to push back ill-defined problems.

Recognising that blockchain is a compound system to onboard, solutions should justify the need among players to leverage a blockchain-enabled platform considering its main characteristics: transparency, privacy, asset ownership, traceability, immutability, trust and decentralisation. ROI targets can help define realistic expectations on all sides. In doing so, a common understanding between industry experts and blockchain engineers is crucial to reveal the real business value derived from using blockchain.

To assess the suitability of blockchain, managers can rely on a decision tree tool [78] and the blockchain model Canvas [101] while keeping abreast of changes in the environment which might imply changes on the latter. The objective of the exercise is, in short, to respond to the question "why a blockchain-enabled platform?". A feasibility study might generate valuable insights into the latter question before moving forward.

Last but not least, incumbents can also consider organising a hackathon, essentially outsourcing the ideation and prototyping phases in a short event.

- **Strategy**

Building upon the initial idea, the founding team must first and foremost define a strategy. An important consideration when defining a blockchain-enabled platform strategy is whether to join an existing network or to build one. Either way, managers are advised to assess what kind of business may be cannibalised in the process.

Disruptive innovation theory recommends startups to design highly differentiated business models and incumbents to create new business units to explore new business models and foster innovation. Incumbents transitioning to a platform business model face specifi challenges that relate to the existing culture, norms and behaviour: change managers will be required to handle the challenge of contrasting organisational identities and business models.

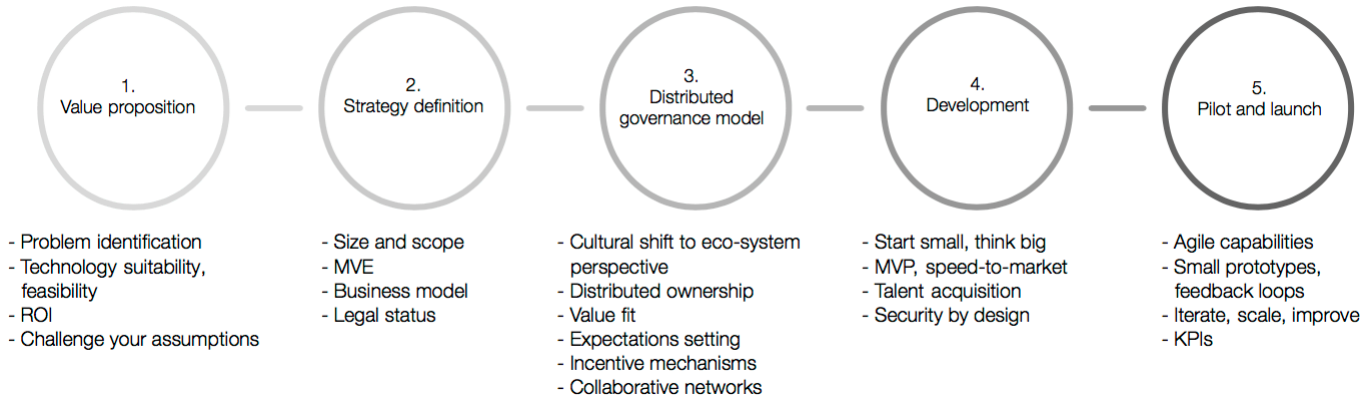


Fig. 2. Key success factors - Blockchain-enabled platforms. Source: autor based on research.

Both startups and incumbents are recommended to rely on the expertise of BaaS and consultancy firms as these play an essential role in supporting entrepreneurs and incumbents on technical to strategic decisions.

- **Size and scope**

The strategy should consider early on the size and scope of the ecosystem to be created. Although blockchain-enabled platforms have the potential to scale globally, this may not be an objective. Managers are actually recommended to start operating in a controlled market and consider whether scalability is an objective right from the beginning, as this influences significantly technological choices. In this phase, managers need to define the desired requirements for the solution taking into account how these differ from an end-user or a business perspective. These requirements should be listed and organised by order of priority.

In starting in a controlled market, blockchain-enabled platforms developers are advised to refrain from radical positions on the decentralized debate and focus on developing commercial success. Once this is the case, migrating to a decentralized model and hence a much bigger ecosystem, can be considered. In short, attempting to go from a prototype to a global solution is too risky to be recommended, as the rate of failure of blockchain projects indicate.

Defining the MVE and the type of stakeholders expected in the ecosystem is the next part the strategy should cover, from end-users to investors, developers, volunteers, curators, moderators, validators, miners, etc. An important consideration is to assess the level of readiness of the different stakeholders, as these may introduce bottlenecks in the development or operation of the platform.

- **Governance**

Relatively quickly after the ideation process, the founding team should define the platform governance model. Mapping the transactions and disputes to arise between the stakeholders identified in the strategy is essential. On-chain processes, like consensus algorithms, are obvious choices to make early on depending on the strategy choices. Off-chain processes are equally important, most notably in permissionless blockchains as these introduce a form of centralisation.

In defining the governance models, blockchain-enabled platform managers may want to consider the legal status of the company behind the platform, and whether a separate entity such as a foundation might be useful or not.

For incumbents, it is important to recognise that although one company may have built the initial platform, for the ecosystem to use the platform, ownership of the latter may be distributed across stakeholders. In order to build collaborative capabilities and retain a sustainable competitive advantage, a cultural shift is required from the silo perspective into an ecosystem perspective. A formal collaboration agreement can secure stakeholder commitment.

Expectation setting is crucial as participants are expected to agree on the provision of resources to build the platform. The founding team would set-up dedicated teams with a driving spirit to make collaboration successful. Without a formal hierarchy, these "self-managed" teams would need to pay great attention to pitfalls of group dynamics such as peer pressure, fear of judgement and groupthink. Diversity being important to spark creativity, the inclusion of minorities such as women and non-technical professionals can contribute to diversify the blockchain ecosystem and therefore reduce the risk of group bias [59].

A decentralised governance model should take into account the following: participants in decentralised platforms tend to place great importance to the value fit with the platform. Managers should beware of potential conflicts of interest that may arise between the platform, its users and those who initiated the project.

- **Incentivization strategies**

Defining incentives for all stakeholders of the ecosystem is critical in ensuring the long-term success of the latter. Permissioned blockchain-enabled platform managers may consider the advantages of traditional seed funding before jumping into tokenization and ICOs, notably when computer illiterate customers are part of the stakeholders.

Tokenization drivers should stem from the governance model and platform objectives and be built into the platform's rules. There is no simple straightforward formula, aside from the fact that constituents need to be involved early on in the incentivization strategy.

Incentive mechanisms should also reward third-party volunteers in critical functions. Notably, there should be incentives for curators, as the ability for decentralised platforms to enhance relevance of listings is key to the user experience and competition with giants like Amazon.

When approaching investors or clients, emphasis should be placed on the value of the solution and not on the underlying technology. However, from an end-user perspective, some solutions would require users to hold tokens and be familiar with the blockchain. The development of educational initiatives is advised to increase user engagement in projects. Teaser Apps can bring tokenization and blockchain closer to people's lives.

Last, a well-rounded communication strategy can help incumbents position their solution, but most importantly build community support. Leveraging word-of-mouth right from the start, by being transparent on values and objectives, proved to be a successful communication tactic for startups.

2. Development Phase

- **Collaborative networks**

Incumbents are recommended to foster innovation outside their own innovation labs. Intercompany collaboration allows to explore new ways of value creation. A key advantage for incumbents is their customer base, trust and credibility. However, incumbents are recommended not to overestimate the innovation efforts and perceive innovation as a journey not a destination. To that respect, disruptive theory recommends not to focus on the short-term results achieved [28], and incumbents are therefore particularly encouraged to assign to those projects a budget independent from short term ROI.

Top management support also appeared as a key ingredient and particularly relevant for large organisations to build an army of volunteers within the company that would advocate for a cultural change and reduce resistance.

- **Finding and retaining talent**

Blockchain development skills are scarce. Whilst incumbents can rely on traditional extrinsic motivators to attract and retain talent, from benefits to mentoring programmes and trainings, startups may use other arguments like equity and impact. Strong emphasis can be put on the mission pursued by the startup and leverage the importance that millennials, a large portion of the blockchain development workforce, give to the cultural fit [91].

Given the scarcity of talent, both startups and incumbents may consider relying on collaborative networks beyond company and industry boundaries. Leveraging global communities of developers can be made possible with the right incentivization strategies.

Last but not least, third party contractors and BaaS companies offer on-demand expertise, although at higher premium.

- **Product development**

Managers are advised to release the product iteratively. For most entrepreneurs, speed-to-market is deemed more important than having the minimum characteristics for the desired solution, even if it means compromising on the research. There are major drawbacks to this approach:

- first, speed-to-market is dependent upon partners' readiness, an issue considering heterogeneous datasets and interoperability costs.
- second, speed-to-market can prove to be a limitation to securing the code. Security considerations are to be discussed early on, notably when it comes to the sensitivity or personal nature of the data to be stored on the blockchain.
- third, the pace of development of the technology calls for flexibility in the coding practices, so as for developers to avoid spending most of their time rewriting their code when the underlying technology changes.

Although it is indeed recommended to develop blockchain-enabled platforms in an agile, iterative process, releasing feature after feature and building the customer base gradually, this paper suggests that the research phase should not be compressed.

- **Securing by design**

Trust is essential in blockchain. Security breaches have profound consequences. As a result, both incumbents and startups should avoid storing sensitive data on any blockchain. The regulatory obligations are just too unclear at the moment to take any chance, given the immutable nature of distributed ledgers.

Startups in particular should pay close attention to their operational security practices, most particularly in terms of credentials management. Defining early on a risk management and a contingency plan, including forks, is also recommended.

Last but not least, given the transparent nature of blockchain code and in particular smart contracts, it is of utmost importance for developers to procure third-party code-auditing for vulnerabilities.

3. Pilot and Launch Phases

- **Regulatory considerations**

Before launching a blockchain-enabled platform in a given market, it is strongly recommended to seek legal advice. National legislation on crypto-projects differs greatly [102]. As such, it may be worth launching the platform in a given jurisdiction where the local authorities are supportive of the project, involving them into the project might be a way to achieve this.

- **Testing, validation and improvement**

Managers are advised to test small prototypes with a subset of customers within a controlled market. In order to enhance customer experience and satisfy evolving customer requirements, managers are highly encouraged to create feedback loops as suggested by the Lean Startup approach [51].

Last but not least, managers are advised to monitor the developments and develop a set of KPIs to monitor platform growth. Metcalfe's law can support on the definition of key performance indicators, in light of the ROI and development timeline previously defined.

VI. Conclusion

Blockchain-enabled platforms have the potential to become no less than the guardians of a post-capitalist world order. The path towards such a bold vision is filled with obstacles, some of which are revealed in this paper. Despite the numerous efforts from scholars and practitioners to demystify blockchain, it remains obscure for most, even those developing blockchain-enabled platforms. Part of this stems from the diverging nature of the technology at the time of writing, leading the vast majority of blockchain projects to fail. Whilst technological convergence will come with time and commercial successes, the failure rate is amplified by a myriad of factors that can actually be tackled today.

Managing expectations is paramount given the vision stated above. Doing so requires a focus on the value proposition of the platform, rather than on the underlying technology. Defining the strategy, governance, size and incentives for the platform is also important before rushing into development. Following this approach should entice startups pursuing disruption through scalability, to paradoxically control growth, and even to make temporary concessions on key blockchain characteristics such as decentralization. Indeed, it is highly probable that the ecosystem stakeholders are simply not ready to operate in a fully decentralized model. Incumbents, on the other hand, need to reconcile with the larger implications of blockchain and embrace new management practices applicable to distributed ecosystems. If not managed properly, the risks are also notable for incumbents, as blockchain-enabled platforms introduce new ways of doing business that can question, if not cannibalize, existing ones.

Actually, developing a blockchain-enabled platform brings about a set of challenges stemming notably from the technology and the skills shortage. Collaborative networks are particularly useful yet difficult to apprehend for traditional actors when it comes to software development, as conflicting interests may arise. Retaining talent in such competitive market can also prove difficult, and open communication on a project's objectives and team values can be determining factors in ensuring a stable workforce in the development phase. Last but not least, uncertainty around regulation calls for caution when launching blockchain-enabled platforms and partnering with well-established stakeholders can reassure investors and regulators.

As the technology progresses and converges, taxonomies will emerge, and further research will shed light on key factors leading to commercial success. With time, managing bias will also be easier, as the enthusiasm depicted by most interviewees certainly exacerbated their perception of the potential and challenges of blockchain-enabled platforms. Particular areas of interest for research will be the actual implementation of decentralized governance models through on and off-chain mechanisms, balancing transparency and privacy, and leveraging collaborative dynamics for the development of blockchain-enabled platforms.

This paper and future research will help clear the path for blockchain-enabled platforms to become not only a reality, but a better one.

Appendix

A. Appendix A: Data Collection

Period	April 2018 - February 2019
Data sources	20 semi-structured interviews and conferences
Interviews Type of organisations	Incumbents: 4 BaaS: 3 Academia: 1 Startups: 6 Consultants: 6
Gender	60% males, 40% females
Job titles	Professor FinTech Professor Founder/Partner CEO Head of Research and Development Blockchain engineer Senior Regulatory Officer CTO Head of Sales Product Manager Data Scientist Head of Innovation and Digital Tranformation Director of Partnerships
Location	US (5), UK (4), Belgium (2), Netherlands (2), Spain (2), Austria (1), Canada (1), France (1), Germany (1) and Switzerland (1).

B. Appendix B: Blockchain-Enabled Platforms Challenges and Recommendations

1. Research Phase

Challenge	Recommendations
The innovation potential of blockchain is debated.	• Startups and incumbents have different views on the innovative nature of blockchain-enabled platforms from moving foward. • Disruptive innovation recommends startups to design highly differentiated business models; and incumbents to create separate business units to explore and drive innovation.
Misguided and misguiding managers lead many platforms to fail.	• Have management define objetives and ROI timeline precisely before jumping into development phase. • Assess wether blockchain is the most suitable technology for the problem at hand. • Learn to say no; push back misguided requests from management. • Do a feasibility study for instance relying on a consultancy firm. • Start with a proof-of-concept and iterate development. • Consider organising a hackathon.

2. Development Phase

Challenge	Recommendations
Blockchain-enabled platform face new and unique governance issues.	• Define governance model before developing technology stack, and no just consensus algorithms. • Map the types of stakeholders expected in the ecosystem and the types of transactions and disputes that will arise. • Consider on and off-chain processes, notably for platform development decisions and for dispute resolution. • Consider the legal status of the company behind the platform, and wether a separate entity such as a foundation might be useful for governance purposes. • For incumbents, accept that although one company may have built the initial platform, for the ecosystem to use tha plataform, ownership of the latter may be distributed across stakeholders.
The skills required to develop blockchain-enabled platforms are diverse, scarce, and under stress.	• Startups should communicate openly about their objetives, values and culture, so as to attract talent who value fit more than benefits. • Incumbents can increase talent pool by creating a mentorship programme and bringing in traditional developers to be mentored by blockchain experts. • Startup and incumbents can rely on collaborative networks beyond company and industry boundaries. • Third-party contractors and Blockchain-as-a-service companies offer on-demand expertise, although at higher premium.
The security implications of blockchain are largely overlooked.	• Avoid at all cost storing sensitive data on blockchain. • Develop early on operational security procedures, notably in terms of credential management. • Develop early on a risk management plan, defining the risk profile and potential threats. • Discuss contingency planning measures, including forks, early on and with ecosystem stakeholders. • Consider seeking advice from security profesionals and companies. • Have third-party code reviews before any launch. • Seek security certification of staff, product and company.
Blockchain technology is not ready to deliver fully on its key promise: decentralization.	• Consider migration costs. • Consider heterogeneity of source data. • Consider level of readiness of ecosystem stakeholders. • Consider the impact of hardware equipment and cloud hosting providers on descentralization. • Consider making temporary concessions by introducing centralised components.

3. Pilot and Launch Phases

Challenge	Recommendations
The mass adoption of blockchain-enabled platforms is constrained by exogenous challenges such as regulatory divergence and access to vital financial services.	• Seek legal advice. • Start small, possibly within a confined jurisdiction. • Consider involving national authorities in project. • Consider registering platform in a country where regulation and financial system is supportive. • Plan early for actual payment processes for operational expenditures.
The mass adoption of blockchain-enabled platforms is contrained by endogenous challenges such as crypto volatility and digital divide.	• Allow users in a platform to redeern tokens with flat currencies to make it more usable and last, to diversify crypto portfolios with asset-backed securities and hedge the market with stablecoins to minimise the effects of crypto price volatility and ultimately protect the capital raised in the form of cryptos. • Simplify access to platform services; train users; edutainment, etc.
Incentives strategies and mechanisms for the launch, development and operation of blockchain-enabled platforms are still exploratory.	• Consider advantages of traditional seed funding before jumping into tokenization and ICOs, notably when it comes to reassuring traditional customers. • Consider the different incentive strategies for the different types of stakeholders, from early investors to miners, moderators and end-users. • Define a communication strategy and build community support, notably leveraging word-of-mouth at the beginning for startups.

Acknowledgment

I would like to express my gratitude to my husband, children and mother, for their active support throughout my studies. I also wish to thank everyone in the blockchain community who contributed to my research, for their availability and openness. Lastly, I would like to extend my sincere appreciation to Dr. Pinar Ozcan and Dr. Philip Drew for their guidance.

References

[1] Cohen, B., Ernesto Amoros, J. & Lundy, L. (2017) The generative potential of emerging technology to support startups and new ecosystems. Kelley School of Business, Indiana University. Business Horizons, 60: 741-745. [online] Available from: https://doi.org/10.1016/j.bushor.2017.06.004

[2] Reillier, L. C. & Reillier, B. (2017) Platform strategy: how to unlock the power of communities and networks to grow your business. London: Routledge. [online] Available from: https://doi.org/10.4324/9781315598949

[3] Puschmann, T. & Alt, R. (2016) Sharing Economy. Business and Information Systems Engineering, 58 (1): 93-99. [online] Available from: https://doi.org/10.1007/s12599-015-0420-2

[4] Friedman, M. (1999) Milton Friedman predicts the rise of Bitcoin in 1999! Youtube. [online] Available from: https://www.youtube.com/watch?v=6MnQJFEVY7s

[5] Devaney, J. (2017) Blockchain and a Renaissance of the Social Commons. Huffpost. [online] Available from: https://www.huffingtonpost.com/entry/blockchain-and-a-renaissance-of-the-social-commons_us_5a4462c2e4b0d86c803c74f0

[6] Burniske, C. & Tatar, J. (2018) Cryptoassets: The Innovative Investor's Guide to Bitcoin and Beyond. McGraw-Hill Education.

[7] Swan, M. (2015) Blockchain: Blueprint for a New Economy. Sebastopol, CA: O'Reilly Media.

[8] Atzori, M. (2015) Blockchain Technology and Decentralized Governance: Is the State Still Necessary? [online] Available from: http://dx.doi.org/10.2139/ssrn.2709713

[9] Dalton-Homes, C. (2018) The Vital Role of Trust in Marketing. Script Consulting. [online] Available from: http://scriptconsulting.co.uk/blog/category/all

[10] Schwab, K. (2016) The Fourth Industrial Revolution. World Economic Forum, Geneva: Penguin.

[11] Schwab, K. (2019) A New Architecture for the Fourth Industrial Revolution. Foreign Affairs. [online] Available from: https://www.foreignaffairs.com/articles/world/2019-01-16/globalization-40

[12] James, A. (2018) 92% of blockchain projects have already failed, average lifespan of 1.22 years. Bitcoinst.com. [online] Available from: https://bitcoinist.com/92-blockchain-projects-already-failed-average-lifespan-1-22-years/

[13] Nakamoto, S. (2008) Bitcoin: A Peer-to-Peer Electronic Cash System. [online] Available from: https://bitcoin.org/bitcoin.pdf

[14] Muzammal, M., Qu, Q. & Nasrulin, B. (2019) Renovating blockchain with distributed databases: An open source System. Future Generation Computer Systems, 90: 105-117.

[15] Pinna, A. & Ruttenberg, W. (2016) Distributed Ledger Technologies in Securities Post-Trading Revolution or Evolution? ECB Occasional Paper No. 172. [online] Available from: https://www.ecb.europa.eu/pub/pdf/scpops/ecbop172.en.pdf

[16] Malki, A. & Weiss, M. B. H. (2016) Automating Ex-Post Enforcement for Spectrum Sharing: A New Application for Block-Chain Technology. [online] Available from: https://ssrn.com/abstract=2754111 or http://dx.doi.org/10.2139/ssrn.2754111

[17] Tapscott, D. & Tascott, A. (2016) The Impact of the Blockchain Goes Beyond Financial Services. Harvard Business Review. [online] Available from: https://hbr.org/2016/05/the-impact-of-the-blockchain-goes-beyond-financial-services

[18] Chameau, J.L., Ballhaus, W.F. & Lin, H.S (2014) Emerging and Readily Available Technologies and National Security: A Framework for Addressing Ethical, Legal, and Societal Issues. Washington (DC): National Academies Press (US); 2, Foundational Technologies. [online] Available from: https://www.ncbi.nlm.nih.gov/books/NBK216326/

[19] Iansiti, M., & Lakhani, K. (2017) The truth about blockchain. Harvard Business Review, 95 (1): 118-127. [online] Available from: https://hbr.org/2017/01/the-truth-about-blockchain

[20] MacDonald, T.J., Allen, D.W.E. & Potts, J. (2016) Blockchains and the Boundaries of Self-Organized Economies: Predictions for the Future of Banking. In: Tasca P., Aste T., Pelizzon L., Perony N. (eds) Banking Beyond Banks and Money. New Economic Windows. Springer, Cham. [online] Available from: https://link.springer.com/content/pdf/10.1007%2F978-3-319-42448-4.pdf

[21] Olleros, F. X. & Zhegu, M. (2016) Research Handbook on Digital Transformations. Université du Québec: Montreal. Edward Elgar Publishing Limited

[22] Peters G.W. & Panayi E. (2016) Understanding Modern Banking Ledgers Through Blockchain Technologies: Future of Transaction Processing and Smart Contracts on the Internet of Money. In: Tasca P., Aste T., Pelizzon L., Perony N. (eds) Banking Beyond Banks and Money. New Economic Windows. Springer, Cham. [online] Available from: https://link.springer.com/content/pdf/10.1007%2F978-3-319-42448-4.pdf

[23] Goorha, P. (2017) The Return of 'The Nature of the Firm': The Role of the Blockchain. The Journal of the British Blockchain Association, 1 (1): 1-5. [online] Available from: http://dx.doi.org/10.2139/ssrn.3080696

[24] Coase, R. H. (1937) The Nature of the Firm. Economica, 4 (16): 386-405. [online] Available from: https://doi.org/10.1111/j.1468-0335.1937.tb00002.x

[25] Catalini, C. & Gans, J.S. (2017) Some Simple Economics of the Blockchain. Rotman School of Management Working Paper No. 2874598; MIT Sloan Research Paper 5191 (16). [online] Available from: http://dx.doi.org/10.2139/ssrn.2874598

[26] Schaffers H. (2018) The Relevance of Blockchain for Collaborative Networked Organizations. In: Camarinha-Matos L., Afsarmanesh H., Rezgui Y. (eds) Collaborative Networks of Cognitive Systems. PRO-VE 2018. IFIP Advances in Information and Communication Technology, 534. Springer, Cham. [online] Available from: https://doi.org/10.1007/978-3-319-99127-6_1

[27] Christensen, C. M. (2006) The Ongoing Process of Building a Theory of Disruption, Journal of Product Innovation Management, 23 (1): 39–55. [online] Available from: https://doi.org/10.1111/j.1540-5885.2005.00180.x

[28] Christensen, C. M., Raynor, M. & McDonald, R. (2015) What Is Disruptive Innovation? Harvard Business Review, 93(12): 44–53.

[29] Rogers, E.M. (2003) Diffusion of innovations. 5th edt.

[30] Pickard, J., Angolia, M. & Chou, T.S. (2018). IPV6 diffusion on the Internet reaches a critical point. Journal of Technology, Management, and Applied Engineering, 34: 1-17. [online] Available from: https://www.researchgate.net/publication/323416689_IPV6_diffusion_on_the_Internet_reaches_a_critical_point

[31] Cheah, E. T. & Fry, J. (2015) Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. Economic Letters, 130: 32-36. [online] Available from: https://doi.org/10.1016/j.econlet.2015.02.029

[32] Böhme, R., Christin, N., Edelman, B. & Moore, T. (2015) Bitcoin: Economics, Technology, and Governance. Journal of Economic Perspectives, 29 (2): 213-38. [online] Available from: https://www.aeaweb.org/articles?id=10.1257/jep.29.2.213

[33] Orcutt, M. (2019) In 2019, blockchains will start to become boring. MIT Technology Review. [online] Available from: https://www.technologyreview.com/s/612687/in-2019-blockchains-will-start-to-become-boring/

[34] Gartner (2019) Gartner 2019 Hype Cycle Shows Most Blockchain Technologies Are Still Five to 10 Years Away From Transformational Impact. Gartner. [online] Available from: https://www.gartner.com/en/newsroom/press-releases/2019-10-08-gartner-2019-hype-cycle-shows-most-blockchain-technologies-are-still-five-to-10-years-away-from-transformational-impact

[35] Morecroft, J.D.W. (2015) Chapter 6: The Dynamics of Growth from Diffusion. Strategic Modelling and Business Dynamics + Website: A Feedback Systems Approach. John Wiley & Sons, Incorporated, New York. [online] Available from: https://ebookcentral.proquest.com/lib/warw/reader.action?docID=1895653

[36] Jeyaraj, A. Rottman, J. W & Lacity, M. C. (2006) A review of the predictors, linkages, and biases in IT innovation adoption research. Journal

of information technology, 21: 1-23. [online] Available from: https://doi.org/10.1057/palgrave.jit.2000056

[37] Lee, Y., Kozar, J.A. & Larsen, K.R.T. (2003) The Technology Acceptance Model: Past, Present, and Future. Communications of the Associations for Informations Systems, 12, (50): 752-780. [online] Available from: https://aisel.aisnet.org/cais/vol12/iss1/50

[38] Alstyne, M., Parker, G., & Choudary, S., (2016) Pipelines, Platforms, and the New Rules of Strategy. Harvard Business Review, 94 (4): 54-62.

[39] Altman, E. J. & Tushman, M. L. (2017) Platforms, Open/User Innovation, and Ecosystems: A Strategic Leadership Perspective. Entrepreneurship, Innovation and Platforms, Advances in Strategic Management, Emerald Publishing Limited, 11: 177-207. [online] Available from: https://doi.org/10.1108/S0742-332220170000037007

[40] Adner, R. (2006) Match Your Innovation Strategy to Your Innovation Ecosystem. Harvard Business Review. [online] Available from: http://sjbae.pbworks.com/w/file/fetch/60084211/Adner_2006_HBR.pdf

[41] Hill, L. A., Brandeau, G., Truelove, E. & Lineback, K. (2014) Collective Genius. The art and practice of leading innovation. Harvard Business Review Press. Boston, Massachusetts.

[42] Utterback, J. M. & Suarez, F. F. (1993) Innovation, competition, and industry structure. Research Policy, 2 (1): 1-21. [online] Available from: https://doi.org/10.1016/0048-7333(93)90030-L

[43] MacCormack, A., Forbath, T., Brooks, P. & Kalaher, P. (2007) Innovation through Global Collaboration: A New Source of Competitive Advantage. Harvard Business School Working Paper 07–080. [online] Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.123.1225&rep=rep1&type=pdf

[44] Faems, D., Looy, B. V. & Debackere, K. (2005) Interorganizational Collaboration and Innovation: Toward a Portfolio Approach. The Journal of Product Innovation Management, 22: 238-250. [online] Available from: https://doi.org/10.1111/j.0737-6782.2005.00120.x

[45] Venkatraman N. & Henderson J.C. (2008) Four Vectors of Business Model Innovation: Value Capture in a Network ERA. In: Pantaleo D., Pal N. (eds) From Strategy to Execution. Springer, Berlin, Heidelberg.

[46] Chesbrough, H. W. (2003) The Era of Open Innovation. MIT Sloan Management Review, 44 (3). [online] Available from: https://eclass.uoa.gr/modules/document/file.php/ECON197/Papers%20Strategy/Chesbrough%202003%20The%20Era%20of%20Open%20Innovation.pdf

[47] Osterwalder, A. & Pigneur, Y. (2010) Business Model Generation. A Handbook for Visionaries, Game Changers, and Challengers. Hoboken, New Jersey: John Wiley & Sons.

[48] Rialti, R., Marzi, G., Silic, M. & Ciappei, C. (2018) Ambidextrous organization and agility in big data era: The role of business process management systems. Business Process Management Journal, 24(5): 1091-1109. [online] Available from: https://doi.org/10.1108/BPMJ-07-2017-0210

[49] O'Reilly, C. A. & Tushman, M. L. (2008) Ambidexterity as a dynamic capability: Resolving the innovator's dilemma. Elsevier: Research in Organizational Behavior, 28: 185-206. [online] Available from: https://www.sciencedirect.com/science/article/pii/S0191308508000105

[50] Trimi, S. & Berbegal-Mirabent, J. (2012) Business model innovation in entrepreneurship. International Entrepreneurship and Management Journal, 8: 449-465. [online] Available from: https://doi.org/10.1007/s11365-012-0234-3

[51] Ries, E. (2011) The Lean Startup: How today's entrepreneurs use continuous innovation to create radically successful businesses. New York: Crown Publishing Group.

[52] Anjum, A., Sporny, M. & Sill, A. (2017) Blockchain Standards for Compliance and Trust, in IEEE Cloud Computing, 4 (4): 84-90. [online] Available from: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8066010&isnumber=8065994

[53] Van den Berg, G. & Pietersma, P. (2015) Key Management Models 3rd Ed. Harlow, UK, Pearson Education Limited.

[54] Osterwalder, A., Pigneur, Y. & Tucci, C. L. (2005) Clarifying Business Models: Origins, Present, and Future of the Concept. Communications of the Association for Information Systems: 16 (1). [online] Available from: https://aisel.aisnet.org/cais/vol16/iss1/1

[55] Osterwalder, A. & Pigneur, Y. (2015) Business model report. Strategyzer. [online] Available from: https://assets.strategyzer.com/assets/resources/business-model-report-2015.pdf

[56] Huertas, J. Liu, H. & Robinson, S. (2018) Eximchain: Supply Chain Finance solutions on a secured public, permissioned blockchain hybrid. [online] Available from: https://www.eximchain.com/Whitepaper-Eximchain.pdf

[57] Lewrick, M., Link, P. & Leifer, L. (2018) The Design Thinking Playbook: Mindful digital transformation of teams, products, services, businesses and ecosystems. Wiley: New Jersey.

[58] King, D. & Lawley, S. (2013) Organisational Behaviour. Oxford: OUP

[59] Jan, Z., Third, A., Bachler, M. & Domingue, J. (2018). Peer-reviews on the blockchain. In: RefResh 2018:1st Workshop on Reframing Research, Cologne, Germany. [online] Available from: http://oro.open.ac.uk/58593/1/Peer_reviews_on_the_blockchain.pdf

[60] Macedo, P., & Camarinha-Matos, L., (2017) Value systems alignment analysis in collaborative networked organizations management. Applied Sciences, 7 (12): 1231. [online] Available from: www.mdpi.com/2076-3417/7/12/1231/pdf

[61] Courtois, N.T. (2016) Features or Bugs: The Seven Sins of Current Bitcoin. In: Tasca P., Aste T., Pelizzon L., Perony N. (eds) Banking Beyond Banks and Money. New Economic Windows. Springer, Cham. [online] Available from: https://doi.org/10.1007/978-3-319-42448-4_6

[62] Scott, B. (2016) How Can Cryptocurrency and Blockchain Technology Pay a Role in Building Social and Solidarity Finance?. United Nations Research Institute for Social Development. [online] Available from: http://hdl.handle.net/10419/148750

[63] Yermack, D. (2017) Corporate Governance and Blockchains. Review of Finance, 21: 7-31. [online] Available from: https://doi.org/10.1093/rof/rfw074

[64] Srinivasan, R., Lilien, G. L. & Rangaswamy, A. (2004) First in, First out? The Effects of Network Externalities on Pioneer Survival. Journal of Marketing, 68 (1): 41-58. [online] Available from: http://journals.ama.org/doi/abs/10.1509/jmkg.68.1.41.24026

[65] Dixon, C. (2017) Crypto Tokens: A Breakthrough in Open Network Design. Medium. [online] Available from: https://medium.com/@cdixon/crypto-tokens-a-breakthrough-in-open-network-design-e600975be2ef

[66] Alabi, K. (2017) Digital blockchain networks appear to be following Metcalfe's Law. Electronic Commerce Research and Application. [online] Available from: https://doi.org/10.1016/j.elerap.2017.06.003.

[67] Wheatley, S., Sornette, D., Huber, T., Reppen, M., & Gantner, R. N. (2018) Are Bitcoin Bubbles Predictable? Combining a Generalized Metcalfe's Law and the LPPLS Model. Swiss Finance Institute Research Paper: 18-22. [online] Available from: http://dx.doi.org/10.2139/ssrn.3141050

[68] Katz, Michael L., and Carl Shapiro. 1994. Systems Competition and Network Effects. Journal of Economic Perspectives, 8 (2): 93-115. DOI: 10.1257/jep.8.2.93

[69] Gandal, N. & Halaburda, H. (2016) Can We Predict the Winner in a Market with Network Effects? Competition in Cryptocurrency Market. [online] Available from: https://www.mdpi.com/2073-4336/7/3/16/htm

[70] Evans, D. S. & Schmalensee, R. (2010) Failure to Launch: Critical Mass in Platform Businesses. Review of Network Economics, 9 (4). [online] Available from: https://doi.org/10.2202/1446-9022.1256

[71] Evans, P., Aré, L., Forth, P., Harlé, N. & Portincaso, M. (2016) Thinking outside the blocks: A Strategic Perspective on Blockchain and Digital Tokens. The Boston Consulting Group. [online] Available from: http://media-publications.bcg.com/BCG-Thinking-Outside-the-Blocks-Dec-2016.pdf

[72] Rauchs, M., Glidden, A., Gordon, B., Pieters, G., Recanatini, M. Rostand, F., Vagneur, K. & Zhang, B. (2018) Distributed Ledger Technology Systems: A Conceptual Framework. Cambridge Centre for Alternative Finance. University of Cambridge. [online] Available from: https://www.jbs.cam.ac.uk/fileadmin/user_upload/research/centres/alternative-finance/downloads/2018-10-26-conceptualising-dlt-systems.pdf

[73] Pazaitis, A., De-Filippi, P. and Kostakis, V. (2017) Blockchain and value systems in the sharing economy: The illustrative case of Backfeed. Technological Forecasting and Social Change, 125: 105-115. [online] Available from: https://doi.org/10.1016/j.techfore.2017.05.025

[74] Carson, B. Romanelli, G., Walsh, P. & Whumaev, A. (2018) Blockchain beyond the hype: What is the strategic business value? McKinsey & Company. [online] Available from: https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/blockchain-beyond-the-hype-what-is-the-strategic-business-value

[75] Ray, J. (2019) On sharding blockchains. Github. [online] Available

from: https://github.com/ethereum/wiki/wiki/Sharding-FAQs#how-does-plasma-state-channels-and-other-layer-2-technologies-fit-into-the-trilemma

[76] Liu, C. (2019) Solving the scalability trilemma: collaborative blockchains in 2019. Crypto Briefing. [online] Available from: https://cryptobriefing.com/scalability-trilemma-collaborative-blockchains/

[77] Lin, I.C. & Liao, T.C. (2017) A Survey of Blockchain Security Issues and Challenges. International Journal of Network Security, 19 (5): 653-659. [online] Available from: https://pdfs.semanticscholar.org/f61e/db500c023c4c4ef665bd7ed2423170773340.pdf

[78] Mulligan, C., Scott, J. Z., Warren, S. & Rangaswami, J.P. (2018) Blockchain beyond the hype: A practical framework for business leaders. World Economic Forum and Imperial College London. [online] Available from: http://www3.weforum.org/docs/48423_Whether_Blockchain_WP.pdf

[79] Swissborg (2019) Crypto Market Review 2018. Swissborg. [online] Available from: https://docsend.com/view/9pda6eq

[80] Malgieri, G. & Custers, B. (2017) Pricing privacy – the right to know the value of your personal data. Computer Law and Security Review, 34 (2): 289-303. [online] Available from: https://www.sciencedirect.com/science/article/pii/S0267364917302819

[81] Angelis, J. & Ribeiro da Silva, E. (2016) Blockchain adoption: A value driver perspective. Business Horizons.

[82] Ethereum (2019) Ethereum: Blockchain App Platform. [online] Available from: https://www.ethereum.org

[83] Mustonen-Ollila, E. & Lyytinen, K. (2003) Why organizations adopt information system process innovations: a longitudinal study using Diffusion of Innovation theory. Information Systems Journal, 13: 275-297. [online] Available from: https://doi.org/10.1046/j.1365-2575.2003.00141.x

[84] Biggam, J (2013) Succeeding with your Master's Dissertation, A step-by-step handbook. McGraw Hill Education: Berkshire, England.

[85] Bryman, A. & Bell, E. (2015) Business Research Methods. Oxford: Oxford University Press.

[86] Ozcan, P. (2018) Growing with the Market? How Changing Conditions during Market Growth Affect Interfirm Ties. Strategic Management Journal, 39 (2): 295-328. [online] Available from: https://doi.org/10.1002/smj.2740

[87] Jick, T. D. (1979) Mixing qualitative and quantitative methods: triangulation in action. Administrative Science Quarterly, 24 (4): 602–611. [online] Available from: https://www.jstor.org/stable/2392366?seq=1#page_scan_tab_contents

[88] Bogdan, R. C. & Biklen, S. K. (1997) Qualitative Research for Education. Boston, MA: Allyn & Bacon.

[89] Ozcan, P., Han, S. & Graebner, M. E. (2017) Single cases: the what, why, and how. In: Mir, Raza A. and Jain, Sanjay, (eds.) The Routledge companion to qualitative research in organization studies. New York: Routledge: 92-112. [online] Available from: https://www.worldcat.org/title/routledge-companion-to-qualitative-research-in-organization-studies/oclc/1002303954&referer=brief_results

[90] Eisenhardt, K. M. & Graebner, M. E. (2007) Theory Building from Cases: Opportunities and Challenges. The Academy of Management Journal, 50 (1): 25-32. [online] Available from: http://www.jstor.org/stable/20159839

[91] Gianniris, D. (2018) The Millennial Arrival And The Evolution Of The Modern Workplace. Forbes. [online] Available from: https://www.forbes.com/sites/forbestechcouncil/2018/01/25/the-millennial-arrival-and-the-evolution-of-the-modern-workplace/#2a17259e5a73/

[92] Ross, A. (2018) Will blockchain solve the cyber security skills crisis? Information Age. [online] Available from: https://www.information-age.com/blockchain-cyber-security-skills-crisis-2-123472476/

[93] Burg, J. Murphy, C. & Pétraud, J.P. (2018) Blockchain for International Development: using a learning agenda to address knowledge gaps. MERL Tech.[online] Available from: http://merltech.org/blockchain-for-international-development-using-a-learning-agenda-to-address-knowledge-gaps/

[94] Janus, E. (2018) Is Blockchain just hot air? New study finds zero percent success rate. Bitcoinist, Cryptocurrency news and technology. [online] Available from: https://bitcoinist.com/blockchain-study-zero-percent-success/

[95] Rancea, B. (2019) What is Ethereum Governance? Complete Beginner's Guide. Unblock. [online] Available from: https://unblock.net/what-is-ethereum-governance/

[96] Dyhrberg A.H. (2016) Bitcoin, gold and the dollar–A GARCH volatility analysis. Finance Research Letters, 16: 85-92. [online] Available from https://www.sciencedirect.com/science/article/pii/S1544612315001038

[97] Tasca, P., Hayes, A. & Liu, S., (2018) The evolution of the bitcoin economy: Extracting and analyzing the network of payment relationships. The Journal of Risk Finance, 19 (2): 94-126. [online] Available from: https://doi.org/10.1108/JRF-03-2017-0059

[98] Wilmoth, J. (2019) Almost Lost My Business: Bitcoin Entrepreneurs Sound Off on Banking Struggles. Cnn. [online] Available from: https://www.ccn.com/bitcoin-entrepreneurs-banking-struggles

[99] Blockchain (2019) Blockchain charts. Blockchain. [online] Available from: https://www.blockchain.com/en/charts/my-wallet-n-users

[100] Willms, J. (2018) Report: Despite Price Volatility Blockchain and Crypto Jobs are In Demand. Bitcoin Magazine. [online] Available from: https://bitcoinmagazine.com/articles/report-despite-price-volatility-blockchain-and-crypto-jobs-are-demand/

[101] Chikara, A. (2019) Blockchain Canvas. 3 Pillar Global. [online] Available from: https://www.3pillarglobal.com/wp-content/uploads/2015/11/blockchain-canvas_final.png

[102] Kent, C. and Jarvis, S. (2017) Divergent Regulatory Approaches to Cryptocurrency Offerings: Developments in Canada, the United States, and China. Capital Markets Bulletin. McMillan. [online] Available from: https://mcmillan.ca/Divergent-Regulatory-Approaches-to-Cryptocurrency-Offerings-Developments-in-Canada-the-United-States-and-China

M. Inmaculada García Sáez

Inma García Sáez is a digital transformation professional with over a decade of operational experience in the financial and high-tech industries. She supports social innovation initiatives tackling global challenges, leveraging emerging technologies. Inma holds an MBA awarded with Distinction from Warwick Business School (United Kingdom). She also studied Digital Transformation and Innovation at INSEAD (France) and Business Administration at the University of Málaga (Spain).

LA UNIVERSIDAD
EN INTERNET

Rectorado
Avenida de la Paz, 137
26006 Logroño (La Rioja)
t (+34) 941 21 02 11

www.unir.net