IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 8, 2019, accepted December 18, 2019, date of publication December 24, 2019, date of current version January 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2962017

Immutability and Decentralized Storage: An Analysis of Emerging Threats

FRAN CASINO[1], (Member, IEEE), EUGENIA POLITOU[1], EFTHIMIOS ALEPIS[1], AND CONSTANTINOS PATSAKIS[1,2], (Member, IEEE)

[1]Department of Informatics, University of Piraeus, 185 43 Piraeus, Greece
[2]Information Management Systems Institute of Athena Research Center, 151 25 Marousi, Greece

Corresponding author: Constantinos Patsakis (kpatsak@unipi.gr)

This work was supported by the European Commission under the Horizon 2020 Programme (H2020) through the Project CyberSec4Europe under Grant 830929 and the Project LOCARD under Grant 832735. The content of this article does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the authors.

ABSTRACT The widespread adoption of the new generation of decentralised architectures, leveraged by blockchain and decentralised file storage (DFS) systems, enables a myriad of new applications and opportunities. Nevertheless, their remarkable features, namely auditability, availability and, among all, immutability, do not come without a cost. In this article, we examine blockchain and the most widely used DFS systems and discuss their main challenges and opportunities, with special regard to their immutability and its impact on their GDPR compliance. A description of current and prospective threats is also provided, along with an analysis of the features that each threat exploits. In addition, we discuss several measures to address the identified threats, and we provide a fertile common ground for further research.

INDEX TERMS Blockchain, distibuted filesystems, decentralised systems, IPFS, immutability.

I. INTRODUCTION

Nowadays, a large number of relevant Internet systems are centralised and/or controlled by lobbies, big companies or governments, e.g. domain name resolution, e-mail, social networks and online storage. In this context, users are forced to follow a set of rules –sometimes unclear– to use such systems. Moreover, the architectural constraints of centralised systems (e.g. single point of failure components), even when they have high redundancy, make these systems vulnerable to Denial of Service (DoS) attacks [1], data misuse [2] and exfiltration attacks [3], [4]. In addition to the aforementioned constraints, the parties managing such systems may apply censorship campaigns [5], [6] to control the data flows from internal user information to data blocks at the network level, escaping thereby from the control of the users. Finally, the management of all these data by a single entity and its correlation with other available pieces of information allows for the extraction of knowledge about individuals without themselves being aware of it, and as such it raises many concerns about privacy and manipulation of the public opinion.

To overcome that situation and paired with the recent hype of Distributed Ledger Technologies (DLTs), decentralised architectures are gaining momentum. Nevertheless, decentralised architectures are not novel. The benefits of such architectures were explored in the early '60s to eliminate single point of failure issues as well as to increase the robustness and redundancy of the systems [7].

Even though these days the continuous digitalisation of our daily lives relieves us of many chores and improves the quality of the received services, it leads at the same time to our extreme dependence on ICT systems. Inevitably, this new landscape of extreme digitalisation, when coupled with human frailty and inherent system vulnerabilities, encourages malevolent behaviours such as cybercrime. Yet, while the constant arms race between cybercriminals and security experts yields to the deployment of new advanced methods and tools, one particular recent development has perplexed the defence mechanisms to a great extent: the development of immutable data structures that are stored and shared across privacy-enhanced decentralised peer-2-peer (P2P) networks. Such immutable structures - even though they provide a landscape of interesting features and applications - can be used as the triggering mechanism of sophisticated and resilient malware campaigns [8].

Malicious phenomena in decentralised environments are definitely neither novel nor unique. In fact, they appeared with the introduction of file-sharing platforms which allowed

The associate editor coordinating the review of this manuscript and approving it for publication was Pierluigi Gallo.

users to exchange illegal content easily [9]. Nevertheless, the problems nowadays are far more complex since content immutability prevents illegal and undesired content from being modified or taken down and, consequently, the correction or erasure of data across decentralised networks cannot be guaranteed. In Europe, for instance, the newly enforced EU General Data Protection Regulation (GDPR) which anticipates the ''*Right to be Forgotten*'' (RtbF) to allow the erasure of personal data under certain conditions, highlights the importance to discuss about decentralised data immutability. However, decentralisation by its nature bypasses all regional jurisdictions and, albeit desirable in some scenarios, it adds another layer of complexity in the case of criminal prosecution.

As previously stated, decentralisation nowadays is most commonly associated with blockchains and their first financial application, bitcoin. Arguably, blockchain technology is one of the catalysts for the present-day technological revolution towards cyber-physical systems which blur the lines between the physical, digital and biological worlds (commonly known as the fourth industrial revolution). Moreover, blockchains are not anymore simple ''*distributed and immutable data ledgers*'' as it has been initially stated in [10]; they realise Smart Contracts (SC) which define a set of functions that are executed by a network of mutually distrusting nodes, enabling thus sophisticated computations of the committed transactions. Notably, the adoption of SCs allows blockchains to operate as a decentralised virtual machine in the form of Decentralised Applications (DApps) and opens the door to numerous new blockchain application scenarios [11].

Blockchains, however, despite their desirable features and inherent security, have already suffered numerous attacks[1] as the ceaseless efforts of malware authors to enhance cyber-crime with sophisticated techniques [12] has created a new ''business'' paradigm. In the context of cryptocurrencies, the increasingly growing value of their market motivates adversaries to exploit weaknesses for profit. The work in [13] presents a systematic study on the threats to blockchain's security by examining popular blockchain systems. Bitcoin, the most prevalent cryptocurrency, has been subject to numerous attacks and exploits that have been analysed in the systematic literature review presented in [14]. The review also discusses the current anonymity considerations in bitcoin as well as the privacy threats to its users, and it proposes some countermeasures based on existing privacy-preserving solutions. Although bitcoin has initially attracted a lot of bad publicity due to its use in paying ransom and other illegal activities, payments, either pseudo or fully anonymous, are only the tip of the iceberg considering the wide range of malicious acts that can be performed by the exploitation of blockchain immutable and transparent nature.

Beyond blockchain, there are various other decentralised P2P networks for file storage and sharing - referred to as Decentralised File Storage (DFS) systems - which are increasingly being employed lately to store off-chain content. Yet, these systems face similar challenges as most of them are either based on immutable data objects, a property that relates to their content addressability, or on blockchain platforms, which are by default immutable data ledgers. While currently blockchains might be an integral part of several DFS architectures allowing several ''flavours'' of DFS systems, in some cases they are used just as a payment method. Nonetheless, when taking into account their properties, they all exhibit similar patterns and can be exploited in similar ways. In this work, we explore the various malicious uses of blockchain and DFS systems arising due to their immutable nature to raise awareness about the issues that Law Enforcement Agencies and Computer Security Incident Response Teams (CSIRTs) will have to face. We argue that beyond trying to foster these technologies as much as possible, we should also aim towards finding ways to counter their possible malicious exploitation. In this regard, this work assesses the threats of these technologies and paves the way for future research by introducing the challenges in the field.

II. DECENTRALISED FILE STORAGE

As previously introduced, P2P file-sharing platforms were already in use throughout the '90s. However, due to the lack of participation/motivation in most of the well-known P2P platforms, BitTorrent appears to be used almost exclusively these days. In a sense, BitTorrent could be treated as a precursor of DFS platforms such as the InterPlanetary File System (IPFS) [15]. Nevertheless, although BitTorrent has evolved in the last years, novel DFS platforms implement a set of features (as we will discuss later in this section) that cannot be presently found in BitTorrent.

DFS systems are decentralised P2P networks used for sharing and storing files across peers in a public network. While their decentralised nature matches the one of blockchains, their scalability and, in most cases, their content-addressability has revealed them as the new alternative to traditional blockchain storage. Of particular relevance is the symbiotic relationship between DFS systems and blockchains, since the combination of both technologies extends their application scenarios [11] to off-chain storage and anonymous file-sharing usually managed through SCs. A brief overview of DFS is provided in Figure 1. Off-chain storage, in particular, i.e. storing files in a DFS system while keeping only a file pointer in the blockchain, has been viewed by several blockchain projects as a convenient - albeit not efficient [16]- method to conform to the GDPR erasure obligations imposed by the RtbF.

Despite the widespread adoption and growth of DFS systems,[2] the most widely known to date is the IPFS. IPFS is a distributed P2P system, both a protocol and a network, for storing and sharing data objects. More precisely, IPFS builds a Merkle Directed Acyclic Graph (DAG), which is

[1] https://www.apriorit.com/dev-blog/578-blockchain-attack-vectors

[2] https://golden.com/decentralised-file-storage-projects/

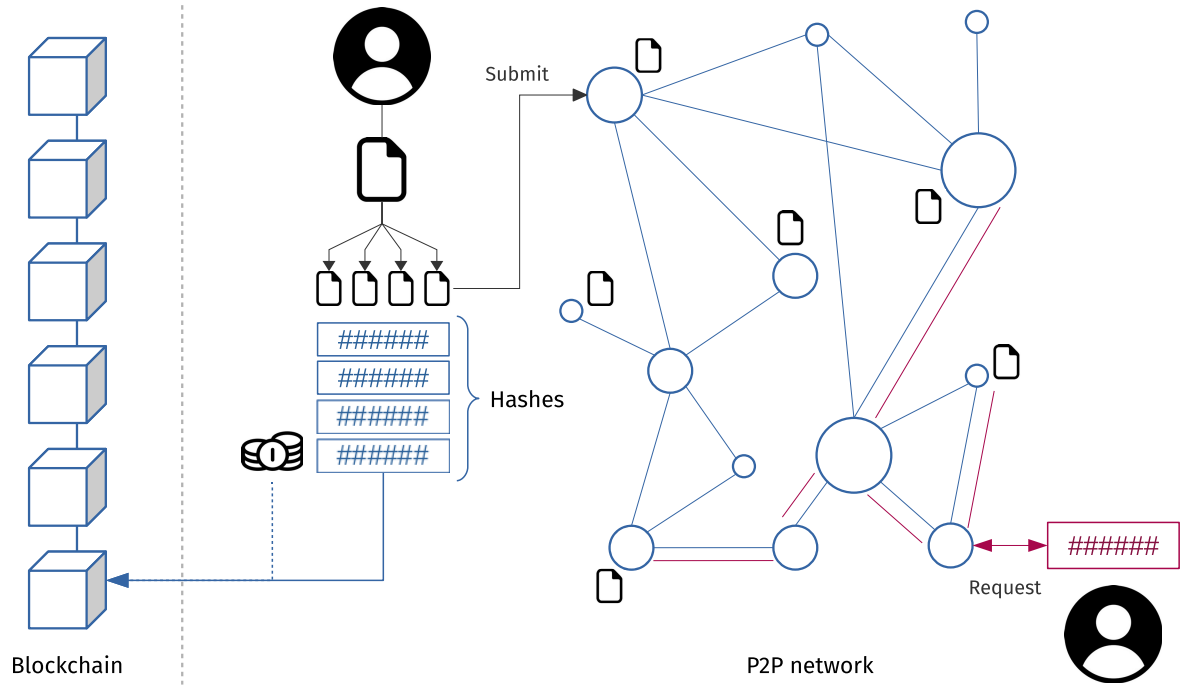

FIGURE 1. An overview of a DFS system. A user splits her files and submits them to the P2P network. The file fragments are distributed all over the network. Once a user wants to retrieve them, a request using the corresponding hashes is made. If a blockchain mechanism exists, a transaction from a user who wants to store a file, pushes the hashes of the files and a monetary deposit to the blockchain.

a cryptographically authenticated data structure, to address data objects. Each object is linked with the hash of its target object and, instead of identifying objects by their location (e.g. HTTPS), the system addresses them by their hash representation of their content (i.e. using Base58 SHA-256) which is by default immutable. While the Merkle DAG structure enables the use of a Version Control System (VCS), the use of the InterPlanetary Naming System (IPNS) and DNSLink protocols allow the creation of mutable addresses to reference always the latest version of an object. Thereby, users can retrieve updated content without knowing the new hashes of such files. Compared with previous P2P file-sharing systems like BitTorrent, IPFS has better deduplication management, since objects are content addressed (by their hash), avoiding duplicates of the exact same content. Moreover, IPNS permits dynamic updating of data pointing to the same address without requiring searching for the new hashes every time a file is updated. On top of that, contents can be accessed in different ways, such as using a terminal or a browser (i.e. in the case of websites). Finally, IPFS implements a proof of storage mechanism, which paired with incentivisation systems like filecoin, opens new market and usability opportunities.

The idea of content addressability, i.e. objects addressed by their content (as opposed to their location), is also exploited by Swarm [17] which operates in the Ethereum service layer. Other well-known examples of DFS are STORJ [18], Maidsafe [19], and Sia.[3] However, not all of these solutions present the same level of maturity to be easily integrated into current production systems. While the IPFS is the most mature among those, since it has already been successfully employed in production by several projects, other DFS platforms such as STORJ, Maidsafe, or the more recent IndImm[4] are still in alpha or beta version or under development. For the sake of clarity and convenience, the main technical characteristics of such systems are summarised in Table 1.

Despite the fact that existing DFS approaches use different architecture implementations, they share a series of common properties/features (with minor exceptions). These common properties are briefly described in Table 2.

III. IMMUTABILITY IN THE DFS ERA

A common characteristic of most DFS systems, as illustrated in Table 2, is their content immutability according to which data held in these systems cannot be tampered with nor be assuredly deleted once they have been uploaded. For some DFS systems (e.g. IPFS) this property stems from their content addressability attribute which specifies that all contents are always addressed by their cryptographic hashes. Hence, a minor edit will always result to a new address (hash) of that file. Furthermore, permanent erasure of data across all IPFS nodes is not guaranteed since as long as there are some nodes sharing the specific content, this will remain available to other peers of the network. Please also note that garbage collection and other deleting options currently provided by the IPFS protocol delete only local "unpinned" copies of shared files. In the case of other blockchain-based DFS systems

[3]https://sia.tech/

[4]https://ndm-inf.github.io/ndm/main

TABLE 1. Technical characteristics of notable DFS systems.

Method	Description	Incentivization	Context of application
IPFS	Decentralised P2P content-based storage. It uses Merkle DAG, DHT Kademlia, BitSwap, IPNS, DNSLink. Proof-of-Retrievability/Proof-of-replication/Proof-of-Spacetime.	Filecoin	Multipurpose
SWARM	Decentralised P2P content-based storage. Integrated with the devp2p multiprotocol network layer of Ethereum blockchain. Proof of Work (to be replaced by Proof of Stake).	Ether	Ethereum ecosystem.
STORJ	Ethereum-based decentralised cloud-based secure storage. Encrypted file sharing, DHT Kademlia, Merkle trees, Sharding, Proof-of-Retrievability.	Storjcoin and others	Multipurpose
Maidsafe	Encrypted data storage and anonymous decentralised file sharing using the SAFE network. Proof-of-Resource.	Safecoin	Multipurpose
SIA	Blockchain-based decentralised file storage platform. It uses Merkle trees, Encryption and Smart Contracts to handle storage requirements. Proof-of-Retrievability/Proof of Work.	SiaCoin	Multipurpose
IndImm	Transaction-based decentralised file storage. Files are chunk into small Base64 string portions and attached through multiple transactions which are chained together sequentially with pointers. Enables search engine which requires knowing the transaction ID, although developers plan to add private transactions by disabling the indexing feature.	XRP	Ripple ecosystem

TABLE 2. DFS main properties.

Property	Description
Content immutability	Content immutability ensures that data cannot be tampered with nor deleted.
Equality	All peers have similar, if not equal, permissions and possibilities.
Decentralisation	The network is totally distributed with no central entities.
Fault tolerance & Attack resilience	A high number of individual peers guarantees the robustness and persistence of the network against attacks such as DDoS.
Availability & Censorship-resistance	The availability of the network depends on multiple peers and not on a single entity. This redundant storage guarantees censorship-resistance.
Unlimited Resources	A high number of simultaneous users sharing their assets.
Scalability	Requests are made to the closest peer instead to a single central location, avoiding bandwidth bottlenecks.
Marketplace monetisation	Incentives for storage are provided through cryptocurrencies. Some coins are still under development (i.e. Filecoin) while others are fully operational (i.e. SiaCoin).

(e.g. SIA, STORJ, SWARM, Maidsafe) in which data are encrypted and split between the network nodes, immutability stems from the inherent blockchain immutable nature. As we have thoroughly discussed in another work [16], blockchains are by default immutable ''append-only'' data ledgers. Blockchain's immutability certifies that transactional data residing in blockchains are tampered-proof, i.e. they can neither be removed nor mutated and as such, they will always remain available in the blockchain network. A recently launched and controversial DFS named IndImm, built on top of Ripple, is advertised as a solution to the immutable uploading and storage of large files. IndImm works by chunking the file into small portions that each fits into a single Ripple transaction and by storing them across multiple transactions which are chained together. Therefore, the immutability is granted by the Ripple's ledger.

Nevertheless, although content immutability may be desired in some contexts, i.e. to ensure censorship resistance, it makes (as discussed later in Section IV-C) DFS systems ideal candidates for malware campaigns as well as illegal/malicious data spreading. Adding to this that most of the DFS systems are endorsed by well-known organisations, such as Cloudfare which supports IPFS, content immutability establishes DFS systems as a fertile ground for wide malware and illegal content dissemination. It also casts doubt on DFS compliance with the European GDPR erasure obligations when individuals, under certain conditions, have the right to request full removal of their personal data held in these networks.

In the IPFS in particular, given that editing a file always produces a new file and hence data modifications are not possible, the impact of immutable IPFS objects has already been under discussion for some time now. In that respect, the IPFS community promotes approaches such as blacklisting to countermeasure the illegal content spreading. However, blacklists present some practical limitations such as their susceptibility to national jurisdictions (i.e. contents may not be illegal in all countries so that blacklists could affect the freedom of speech) as well as their scalability which renders them useless if they grow arbitrary long. Besides, blacklisting may end up not being adopted by all nodes in the network. Moreover, when the size of the blacklisted content grows beyond a given point, the additional overhead is expected to dissuade many nodes from using it. Therefore, popular illegal content is likely to remain on the network as long as there are incentives for nodes to keep it available (see Table 1) or as long as infected devices store and distribute it. Although most of the non-blockchain based DFS systems allow the deletion of data objects stored in a single node, it is worth noting that erasure across the entire network is a rather difficult task. The problem is exacerbated in the cases of IPFS and IndImm if raw data are stored since no global erasure mechanisms are implemented.

Based on the above, there is an urgent need to identify DFS systems' vulnerabilities and privacy limitations arising from their immutability and to discuss appropriate technical measures and prevention mechanisms against their malicious use and towards their alignment with the RtbF.

IV. EMERGING THREATS

In what follows, we attempt to provide an overview of the emerging threats that are becoming relevant due to the adoption of immutable decentralised storage.

A. PERSONAL DATA

Malevolent uses for promoting and disseminating sensitive data content and infringing upon human rights, such as the

rights to privacy and to data protection, can be exploited in both blockchain and DFS systems. Of specific relevance is the erasure obligations of the RtbF enshrined in the Article 17 of the GDPR according to which individuals have the right, when certain conditions apply, to request the erasure of their personal data from all the places to which they have been disseminated [20]. Certainly, such a requirement may have a huge impact on contemporary information systems as well as on future technological developments. In the case of blockchain, in particular, its immutable nature contradicts the RtbF when personal data are at stake. As discussed in previous works, blockchain compliance with the GDPR only through the use of hash values and public-key cryptography cannot be guaranteed since such mechanisms are pseudonymous, not anonymous, and therefore relevant data are not exempted from the GDPR [16]. Nevertheless, blockchains do not have to expose personal data directly to reveal individuals' personal information, since sensitive data (e.g. health status and visiting records) can be also leaked by exploiting metadata information [16]. As a result, when personal data are at stake, namely stored in blockchain transactions, blockchain's immutability contradicts the data protection right imposed by the RtbF since they can never be deleted [16]. In the case of DFS systems like IndImm and IPFS, although they promote the dissemination of all types of data without restriction, they do not support any efficient methods for completely and assuredly removing any personal or sensitive content published across the entire network. By all accounts, the RtbF anticipated by the GDPR imposes a critical challenge in terms of storing personal data in blockchains or in other DFS systems.

B. ILLEGAL CONTENT

As already discussed, blockchains are by default immutable append-only data ledgers. Clearly, this property is a double-edged sword when it comes to using these networks abusively or for malevolent causes. For example, when Bitcoin Satoshi Vision (BSV) [21] data limits on transactions increased to 100KB, users started to store webpages, images, and video in just one transaction. In fact, BSV was recently used to store illegal contents about child abuse [22]. Similarly, in March 2018 researchers in Germany [23] found that of 1,600 files stored on the bitcoin blockchain, two included lists of 274 links to child porn websites (including Tor hidden services). An extensive study about the types of data stored in bitcoin, including, among others, illegal and copyrighted content as well as malware, can be found in [23].

Beyond blockchains, DFS systems are also vulnerable to hosting and disseminating illegal content since data stored in those systems cannot be verifiably deleted across the entire network. Even if their content-addressed feature enables a fast finding of such data, a slight modification of a file can end up with multiple versions of the same data. IndImm, as already discussed, has sparked many discussions not only due to the threat of having illegal content stored immutably on such a network, but also due to the low transaction costs involved in hosting large files since this potential encourages, among others, the blackmailing with illicit content whose indefinite and permanent store is now possible. The problem of storing permanently illegal or malicious content in blockchains and DFS systems is further exacerbated by the fact that current legislation does not deal with decentralised technology at all. Not even the European GDPR has taken into account the immutable nature of decentralised storage systems, let alone blockchain, when legislated the RtbF [20].

C. MALWARE

Blockchains have already been explored for malware exploitation as they can be used to enhance malware spreading and persistence due to its immutability. For instance, a malicious botnet can be managed and coordinated based on transaction information in the bitcoin blockchain [24]. This allows the malware author to update the location of the server dynamically in real-time, and as the malware directly goes to the right location, no longer generates a sequence of NXDomain responses which are one of the primary detection mechanisms. Other examples include the implementation of new viral techniques that leverage the blockchain network [25].

In the case of DFS, the issue is even more complicated, as already described in Section III. The performance and spreading capabilities of DFS networks enable hard-to-takedown malware campaigns, like the one taken place by the IPStorm.[5] More concretely, DFS systems such as IPFS allow a set of opportunities for malware authors: immutable storage, costless deployment, seamless content dissemination, and anonymity. For instance, bots can disseminate the commands of the C&C server only by pinning a file, or a set of files with different versions/variations of a malware, which translates into a resilient and mutable botnet [8]. Moreover, since the IPFS, as well as other DFS systems, is content-addressable, it is impossible to know all the affected files in the network unless either their hash is already known or a search for specific content across the network is performed. The odds for the latter, however, are decreased by the fact that files may be split across different nodes or be in encrypted form. Therefore, as exploited in [8], malware may use *Resource Identifier Generation Algorithms (RIGAs)* to generate arbitrary amounts of requests to the underlying DFS based on the corresponding address, making the quest for finding the precise location of malicious content impractical, as in the case of domain generation algorithms.

D. UPDATE AND DEPRECATION

Software life-cycle, apart from its continuous iterations dictated by almost every software development methodology, it is characterised by two standard milestones: its updates, e.g. due to changes in business or system requirements; and its withdrawals, e.g. due to technological advances that push

[5]https://www.anomali.com/blog/the-interplanetary-storm-new-malware-in-wild-using-interplanetary-file-systems-ipfs-p2p-network

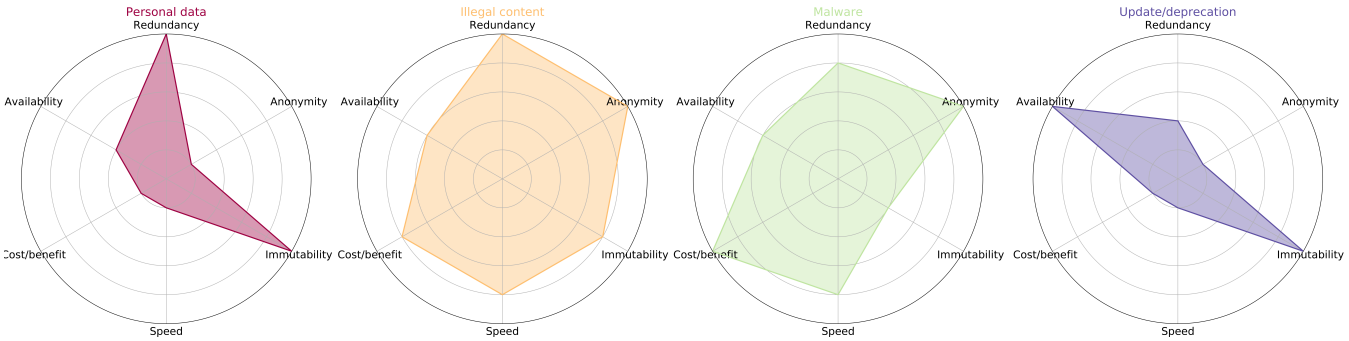

FIGURE 2. DFS and blockchain features exploited by each threat.

toward the development of new solutions, rendering thus the previous ones deprecated.

Nevertheless, decentralised P2P networks like DFS and blockchain require the development and adoption of permanent data structures for which software revisions are not possible, and there is not any expected time-frame for their withdrawal. In fact, while software clients can be updated, this does not apply to other components such as Smart Contracts (SCs) which are deployed once and are expected to run forever. Given the security threats that emerge every now and then, one can assume that already deployed SCs, as well as future ones, could be vulnerable not only to current attacks but to those that are to be disclosed in some years from now. In addition, SCs pointing to deprecated data stored in DFS may hinder this issue to a greater extent. As it has been demonstrated, SCs vulnerabilities pose significant threats to the correct operation of the DApps, especially in terms of monetary losses. In this context, the cascading effect of a potential attack to SCs could be overly detrimental due to their immutable nature.

An evaluation of the impact the different features of blockchain and DFS have on the previously discussed threats is illustrated in Figure 2. For each threat, we illustrate which feature is more relevant and likely to be exploited. In the case of personal data, immutability and redundancy are clearly the most challenging features to overcome, since they prevent their permanent erasure. Illegal content shares similar concerns, with the addition of anonymity and an ideal cost/benefit trade-off since illegal content can be spread almost in real-time and at a very low cost. Malware shares similarities with illegal content, yet the immutability feature here is not so relevant since new versions of malware can easily be created and spread again, as stated in Section II. Finally, the update and deprecation threat can be clearly exploited if data are continuously available and immutable since vulnerable versions of specific software or smart contracts will be permanently accessible.

V. PREVENTIVE MEASURES

Several strategies can be adopted to overcome the challenges posed by content immutability in blockchains and DFS systems.

A. BLOCKCHAINS

In terms of blockchains, technical efforts to circumvent immutability while preserving their inherent security are steadily emerging [16]. Among them, state-of-the-art cryptographic techniques, as well as various common workarounds such as off-chain storage and encryption, have been introduced. Yet, in order to limit any undesired effects of immutability in blockchain environments, some more crucial measures need to be taken by project stakeholders:

1) MINIMAL DATA STORAGE

Data management and storage policies should be defined across the blockchain to ensure at least availability and auditability. In particular, measures such as storing data in an encrypted form, or storing data off-chain while keeping their hashes in the blockchain, or employing role-based access control to manage parties' permissions according to their role, are some indicative examples that can minimise the impact of content immutability.

2) THREAT-RISK ASSESSMENT

The use of blockchain security reference models for studying the nature and hierarchy of vulnerabilities and security threats, their origin as well as mitigation techniques or countermeasures [26] is a desirable feature. The outcome of such a model may help to discover threats at different levels (i.e. network layer, consensus layer, replicated state machine layer, and application layer) for old and novel blockchains, which paired with a proper suitability assessment [11] (i.e. in the case of deploying a blockchain-based application), may diminish future vulnerabilities. Moreover, technical and organisational procedures in case of failures or exploits should be in place. Furthermore, emergency plans and efficient resolution mechanisms will minimise the impact of malware campaigns. This is of high importance in cases of monetary losses where an immediate response is critical.

3) SMART CONTRACTS

Regarding SCs, several additional features should be considered. For instance, SC deactivation, using e.g. a boolean available only to its owner, which can be checked at the beginning of each function is a way towards controlling

their termination. Moreover, approaches of having upgradable features such as calls to external SCs [27] may minimise the amount of information stored in the blockchain and the impact of exploits in case of vulnerabilities. Moreover, several works aim at analysing SCs' vulnerabilities to enhance their security. For instance, ContracFuzzer [28] generates fuzzing inputs based on the ABI specifications of SCs, defines test oracles to detect vulnerabilities, and analyses logs of SCs runtime behaviour to detect security flaws. Similarly, Securify [29] analyses Ethereum SCs by extracting the contract's dependency graph to reveal precise semantic information from the code. Next, it checks compliance and violation patterns according to a predefined set of properties. A more general vulnerability analysis can be found in [30] in which authors analyse the security vulnerabilities of Ethereum SCs, providing a taxonomy of common programming pitfalls that may affect their security. Due to the pivotal role of SCs security on the reliability of blockchain transactions, user communities to help in finding and fixing SC vulnerabilities have become an emerging trend lately (https://www.dasp.co/). Finally, a recent summary of the most relevant formalisation techniques used to verify the functional correctness of a smart contract can be found in [31].

B. DFS

Overcoming content immutability's undesired effects in DFS systems is a challenging area of research both for academia and DFS systems' core development teams. However, thus far, there have not been any plans to support mutable data objects in DFS systems. Admittedly, this would not be an easy task as many DFS systems have been built on top of public blockchains for which mutability is not an option. Nevertheless, there exist some DFS, such as IPFS and STORJ, which claim that support data deletion. Yet, this deletion is not assured across all the nodes of the network since as long as someone has enough incentives to keep data alive, data will not be deleted across all the network nodes nor will be expired. Although this can be partially solved by the use of encryption in systems like STORJ, IPFS does not provide object-level encryption yet. Unavoidably, some sort of mutability or erasure mechanism needs to be integrated into DFS networks in order, on the one hand, to protect them from malicious and illegal content dissemination and, on the other, to make them compliant with the legislated data protection rights such as the RtbF defined in the GDPR.

Nevertheless, the adoption of such measures in a protocol level is currently a controversial topic. An argument against this proposal is the fact that a mechanism like the one proposed would be mostly useless to protect against specific malicious behaviours. For instance, files containing malware or personal data can always be slightly modified and re-uploaded without their content to be significantly affected. In that regard, user's participation and behaviour should be rewarded appropriately, not only to prevent such malicious behaviour but to promote the continuous use of the decentralised systems preventing this way the lack of motivation/participation in these systems, which - as discussed in Section 1- happened in the past with other P2P systems. This is particularly relevant in the case of DFS in which users *lent* their resources (mainly their local data storage) to be used by the network. While for the time being proper rewarding mechanisms include mainly cryptocurrency payments, the incorporation of advanced identity management standards (e.g. ERC725 and ERC735 in Ethereum and the Self-Sovereign Identity (SSI) concept [32], [33]) may extend this reward mechanism to more specialised and personalised services. In parallel, efforts should be devoted to the design of mechanisms for reducing the impact of users leaving the system, as this kind of behaviour may lead to unrecoverable data loss incidents or even to the complete take-down of the system.

VI. CONCLUSION

In this paper we analysed the landscape of the emerging threats for the next generation decentralised systems, focusing mainly on file storage systems and blockchain. We identified four major threats relating to personal data, legal content, malware, and update/deprecation and we highlighted the relative weights with which several - otherwise desirable - features of the decentralised systems (e.g. availability, anonymity, redundancy) may contribute to each threat. Moreover, we discussed in detail the challenges and possible countermeasures in each case, providing a fertile ground for further research.

Future work will focus on exploring the existing threats more thoroughly (e.g. we consider that further awareness should be raised about the lack of motivation in these systems which - as already shown in the past with some P2P systems - can be critical to the protection of their data). Efforts also will be made for identifying new threats since the capabilities of the next generation decentralised systems are yet to be fully exploited.

REFERENCES

[1] *The Largest DDOS Attacks of All Time*. Accessed: Oct. 2019. [Online]. Available: https://www.cloudflare.com/learning/ddos/famous-ddos-attacks/

[2] M. Foulsham, "Living with the new general data protection regulation (GDPR)," in *Financial Compliance*. Cham, Switzerland: Springer, 2019, pp. 113–136.

[3] S. Mohurle and M. Patil, "A brief study of wannacry threat: Ransomware attack 2017," *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 5, pp. 1–4, 2017.

[4] *Data Breaches 101: How they Happen, what Gets Stolen, and where it All Goes*. Accessed: Oct. 2019. [Online]. Available: https://www.trendmicro.com/vinfo/us/security/news/cyber-attacks/data-breach-101

[5] D. Bamman, B. O'connor, and N. Smith, "Censorship and deletion practices in Chinese social media," *First Monday*, vol. 17, no. 3, Mar. 2012.

[6] J. Pan, "How the market for social media shapes strategies of Internet censorship," in *Digital Media and Democratic Futures*. Philadelphia, PA, USA: Univ. of Pennsylvania Press, 2019, p. 196.

[7] P. Baran, "On distributed communications networks," *IEEE Trans. Commun. Syst.*, vol. CS-12, no. 1, pp. 1–9, Mar. 1964.

[8] C. Patsakis and F. Casino, "Hydras and IPFS: A decentralised playground for malware," *Int. J. Inf. Secur.*, vol. 18, no. 6, pp. 787–799, Dec. 2019.

[9] T. Klumpp, "File sharing, network architecture, and copyright enforcement: An overview," *Managerial Decis. Econ.*, vol. 35, no. 7, pp. 444–459, Oct. 2014.

[10] S. Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System*. 2008.
[11] F. Casino, T. K. Dasaklis, and C. Patsakis, "A systematic literature review of blockchain-based applications: Current status, classification and open issues," *Telematics Inform.*, vol. 36, pp. 55–81, Mar. 2019.
[12] S. Mansfield-Devine, "The malware arms race," *Comput. Fraud Secur.*, vol. 2018, no. 2, pp. 15–20, Feb. 2018.
[13] X. Li, P. Jiang, T. Chen, X. Luo, and Q. Wen, "A survey on the security of blockchain systems," *Future Gener. Comput. Syst.*, to be published.
[14] M. Conti, E. S. Kumar, C. Lal, and S. Ruj, "A survey on security and privacy issues of bitcoin," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3416–3452, 4th Quart. 2018.
[15] J. Benet, "IPFS-content addressed, versioned, P2P file system," 2014, *arXiv:1407.3561*. [Online]. Available: https://arxiv.org/abs/1407.3561
[16] E. Politou, F. Casino, E. Alepis, and C. Patsakis, "Blockchain mutability: Challenges and proposed solutions," *IEEE Trans. Emerg. Topics Comput.*, to be published.
[17] V. Tron. (2016). *Swarm*. [Online]. Available: https://swarm-gateways.net/bzz:/theswarm.eth/#the-thsph-orange-paper-series
[18] B. Produit. (2018). *Using Blockchain Technology in Distributed Storage Systems*. [Online]. Available: https://courses.cs.ut.ee/MTAT.07.022/2018_spring/uploads/Main/bruno-report-s17-18.pdf
[19] *Maidsafe*. Accessed: Nov. 8, 2019. [Online]. Available: https://maidsafe.net/
[20] E. Politou, E. Alepis, and C. Patsakis, "Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions," *J. Cybersecur.*, vol. 4, no. 1, p. tyy001, Jan. 2018.
[21] *Bitcoin Satoshi Vision*. Accessed: Nov. 8, 2019. [Online]. Available: https://bitcoinsv.io/
[22] (2019). *Child Abuse Images Hidden in Crypto-Currency Blockchain*. [Online]. Available: https://www.bbc.com/news/technology-47130268?ocid=socialflow_twitter
[23] R. Matzutt, J. Hiller, M. Henze, J. H. Ziegeldorf, D. Müllmann, O. Hohlfeld, and K. Wehrle, "A quantitative analysis of the impact of arbitrary blockchain content on bitcoin," in *Proc. 22nd Int. Conf. Financial Cryptogr. Data Secur. (FC)*. Berlin, Germany: Springer, 2018.
[24] S. Pletinckx, C. Trap, and C. Doerr, "Malware coordination using the blockchain: An analysis of the cerber ransomware," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, May 2018, pp. 1–9.
[25] J. Moubarak, M. Chamoun, and E. Iliol, "Developing a k-ary malware using blockchain," in *Proc. IEEE/IFIP Netw. Oper. Manage. Symp. (NOMS)*, Apr. 2018, pp. 1–4.
[26] I. Homoliak, S. Venugopalan, Q. Hum, D. Reijsbergen, R. Schumi, and P. Szalachowski, "The security reference architecture for blockchains: Towards a standardized model for studying vulnerabilities, threats, and defenses," 2019, *arXiv:1910.09775*. [Online]. Available: https://arxiv.org/abs/1910.09775
[27] I. Sergey and A. Hobor, "A concurrent perspective on smart contracts," in *Proc. Int. Conf. Financial Cryptogr. Data Secur.* Cham, Switzerland: Springer, 2017, pp. 478–493.
[28] B. Jiang, Y. Liu, and W. K. Chan, "Contractfuzzer: Fuzzing smart contracts for vulnerability detection," in *Proc. 33rd ACM/IEEE Int. Conf. Automat. Softw. Eng. (ASE)*, New York, NY, USA, 2018, pp. 259–269.
[29] P. Tsankov, A. Dan, D. Drachsler-Cohen, A. Gervais, F. Bünzli, and M. Vechev, "Securify: Practical security analysis of smart contracts," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, New York, NY, USA, 2018, pp. 67–82.
[30] N. Atzei, M. Bartoletti, and T. Cimoli, "A survey of attacks on ethereum smart contracts (sok)," in *Principles of Security and Trust*, M. Maffei and M. Ryan, Eds. Berlin, Germany: Springer, 2017, pp. 164–186.
[31] A. Singh, R. M. Parizi, Q. Zhang, K.-K.-R. Choo, and A. Dehghantanha, "Blockchain smart contracts formalization: Approaches and challenges to address vulnerabilities," *Comput. Secur.*, vol. 88, Jan. 2020, Art. no. 101654.
[32] M. S. Ferdous, F. Chowdhury, and M. O. Alassafi, "In search of self–sovereign identity leveraging blockchain technology," *IEEE Access*, vol. 7, pp. 103059–103079, 2019.
[33] A. Mühle, A. Grüner, T. Gayvoronskaya, and C. Meinel, "A survey on essential components of a self-sovereign identity," *Comput. Sci. Rev.*, vol. 30, pp. 80–86, Nov. 2018.

FRAN CASINO was born in Tarragona, Spain, in 1986. He received the B.Sc. degree in computer science, the M.Sc. degree in computer security and intelligent systems, and the Ph.D. degree (Hons.) *(cum laude)* in computer science from Rovira i Virgili University, Tarragona, Catalonia, in 2010, 2013, and 2017, respectively. He was a Visiting Researcher with ISCTE-IUL, Lisbon, in 2016. He has participated in several European-, Spanish-, and Catalan-funded research projects. He is currently a Postdoctoral Researcher with the Department of Informatics, Piraeus University, Piraeus, Greece. His researches focus on pattern recognition, and data management applied to different fields, such as privacy and security protection, recommender systems, smart health, and blockchain. He received the Best Dissertation Award during his Ph.D. degree.

EUGENIA POLITOU received the Diploma (B.S.E.) and M.Sc. degrees in electrical and computer engineering from the Democritus University of Thrace, Xanthi, Greece. She is currently pursuing the Ph.D. degree in informatics with the University of Piraeus, Greece. Her current research interests include privacy and data protection in decentralized networks and other state-of-the-art technologies, such as mobile computing. She has a long experience in research, security, analysis, and system design under various national and European large-scale IT projects within the private and public sector. She also works as an Information Security Officer with the Independent Authority for Public Revenue, Greece.

EFTHIMIOS ALEPIS received the B.Sc. degree in informatics and the Ph.D. degree from the Department of Informatics, University of Piraeus, Greece, in 2002 and 2009, respectively. He is currently an Assistant Professor with the Department of Informatics, University of Piraeus, since December 2013. He has authored/coauthored more than 60 scientific articles which have been published in international journals, book chapters, and international conferences. His current research interests are in the areas of object-oriented programming, mobile software engineering, human–computer interaction, affective computing, user modeling, and educational software.

CONSTANTINOS PATSAKIS received the B.Sc. degree in mathematics from the University of Athens, Greece, the M.Sc. degree in information security from Royal Holloway, University of London, and the Ph.D. degree in cryptography and malware from the Department of Informatics, University of Piraeus. He has authored more than 100 publications in peer reviewed international conferences and journals. He has participated in several National and European Research and Development projects. Additionally, he has worked as a Researcher with the UNESCO Chair in data privacy and as a Research Fellow with Trinity College. He is currently an Assistant Professor with the University of Piraeus and an Adjunct Researcher of the Athena Research and Innovation Center. His main areas of research include cryptography, security, privacy, data anonymization, and data mining.