

# Developing a K-ary malware using Blockchain

Joanna Moubarak  
*ESIB, USJ*  
*CIMTI*  
Beirut, Lebanon  
joanna.moubarak@net.usj.edu.lb

Eric Filiol  
*ESIEA*  
*(C + V)<sup>O</sup> Lab*  
Laval, France  
efiliol@netc.fr

Maroun Chamoun  
*ESIB, USJ*  
*CIMTI*  
Beirut, Lebanon  
maroun.chamoun@usj.edu.lb

**Abstract**—Cyberattacks are nowadays moving rapidly. They are customized, multi-vector, staged in multiple flows and targeted. Moreover, new hacking playgrounds appeared to reach mobile network, modern architectures and smart cities. For that purpose, malware use different entry points and plug-ins. In addition, they are now deploying several techniques for obfuscation, camouflage and analysis resistance. On the other hand, antiviral protections are positioning innovative approaches exposing malicious indicators and anomalies, revealing assumptions of the limitations of the anti-antiviral mechanisms. Primarily, this paper exposes a state of art in computer virology and then introduces a new concept to create undetectable malware based on the blockchain technology. It summarizes techniques adopted by malicious software to avoid functionalities implemented for viral detection and presents the implementation of new viral techniques that leverage the blockchain network.

**Index Terms**—Malware, K-ary Virus, Malicious program, Blockchain, APT

## I. INTRODUCTION

Computer infections hit the mainstream in recent years exploiting systems vulnerabilities and creating specific malicious software that are penetrating organizations and governments for damage purposes or to steal information. Also, with the emerging new services and technologies, the marketplace has fully-fledged to reach the cloud, the Internet of Things (IoT) and the interconnected world adding new proliferation environments for malware and augmenting the viral infection risk which is assessed depending on the number of infections and their impacts, the detection ability, the protection in place and the capacity to disinfect and isolate a convicted system.

In 1982, a boot virus was introduced [1]. Also, the first official virus appeared in 1983 [1] under UNIX proving that no operating system is immune against vulnerabilities. In 1988, viruses and worms leave the laboratories. Since then, malware evolution is exponential. The first viruses and worms for mobile phones appeared in 2004 [2]. In 2008, Stuxnet malware marked a turning point in the enhancement and professionalization of the attacks and the damage they can engender [3], [4]. Advanced Persistent Threats (APTs) became more omnipresent from 2015, after the hack of Carbanak [5] and Sony Pictures [6] discovered in 2014. In 2016, the Locky ransomware hit considerably several places mainly in Europe [7]. Also, the Mirai DDoS attacks on IoT made headlines and derived motives for software developers to incorporate universal security protections [8]. Recently, the WannaCry

ransomware [9] spreads laterally and a new form of the ransomware called Petya targeted several countries worldwide [10]. Furthermore, new Android and iOS malware are discovered each day. These attacks summarize the importance of viral threats and illustrate the evolving viral risk. For instance, the level of risk is potentially high for targeted attacks, probably low to medium for other actors. Furthermore, the main risk today consists in the creation of botnets network utilized to create several types of attacks.

Besides, antiviral industry is constantly enhancing its capabilities to reduce the gap between the detection and networks containment and always striving to mitigate the breach by combining several analysis mechanisms and machine learning algorithms. However, security solutions can only decrease risks without eliminating them. In reality, the attack has more advantages over the defender. The attacker not only always has the initiative but he innovates constantly. As soon a technique is deemed impossible, the attacker will try to bypass it. Moreover, an attacker will usually seek to hide as long as possible what he has managed to do. Furthermore, the attack always has a lead in time [11].

This paper is the result of a prolonged survey on viral techniques adopted by malware as we go through several computer virology studies [12] [13] [14]. Also, some malware samples were examined in real time against several analysis approaches [15] and in the other hand, many antiviral solutions were tested in different attack scenarios and networks. Moreover, several blockchain architecture types [16] were considered while developing the new malware.

The remainder of this paper is organized as follows: Section II develops a summary of viral and antiviral techniques. Section III concentrates on k-ary malware, followed by the utilization of the blockchain potential to create a k-ary malware in Section IV. We conclude this paper and present our future work in Section V.

## II. TECHNICAL ANALYSIS

The threat landscape is getting more complicated and businesses remain agile. Many challenges continue to strive security strategies in order to add expanded detection capabilities, find a solution that fits into the architectures and stay within acceptable levels of operational risk. This section gives an overview on viral techniques employed and exposes how antiviral solution providers address those challenges.

arXiv:1804.01488v1 [cs.CR] 4 Apr 2018

### A. Computer infections

Traditionally, the term computer virus is misused to generally refer to offensive programs [17]. Currently, malicious softwares are categorized depending on several viral mechanisms. L. Adelman [17] divided malware into two disjoint categories: *simple* and *Self-reproducing*. Simple malware may be alienated to logic bombs and trojan horse. This category installs itself either in a resident, stealth or persistent mode. Whereas, self-reproducing malware try to overlap all or parts of its malicious code into another program. This class encompasses viruses and worms. Computer infections, whether simple or self-replicating, are installed on a system to compromise the confidentiality, the integrity or the availability of the system. Additionally, the main methods of propagation rely on file sharing, network exchanges, P2P, emails communications and downloading. Mobile computing is another vector of propagation, including direct LAN-WAN and smart phones connections. Recently, malware are surrounding the IoT technology recruiting intelligent devices as zombies to conduct attacks. The recent evolution of infective programs has shown that the scope has surpassed the computer to reach exotic platforms and that the threat becomes global. The mechanisms can certainly vary for one system to another.

There are several definitions of the concept of computer infection, but none is really complete as recent developments are not taken into account. Attacks by computer infections are all based more or less on social engineering. Another important aspect of the mode of action of the infecting programs is the presence of software vulnerabilities that make the exploitation possible independently of the users.

### B. Antiviral Techniques

Antiviral Techniques can be alienated to the following:

- 1) Static antiviral techniques: These techniques examine the codes without execution.
  - a) Viral signatures: This technique looks for any arrangement of bits and instructions that distinguishes a particular program.
  - b) Spectral analysis: This technique consists in examining the code functions and instructions.
  - c) Heuristic analysis: This technique studies the performance and the behavior of a particular program based on policies and guidelines.
- 2) Dynamic antiviral techniques: These techniques execute the code for analysis.
  - a) Behavior monitoring: Many mechanisms to monitor the related indicators of compromise (IOCs).
  - b) Code emulation: This technique loads the program into a specific memory zone to mimic the code execution.
- 3) File integrity checking: This technique checks any modification in critical files.

Most efficient antiviral solutions are combining several different antiviral techniques to fight against malware [18]–[21]. A layered approach is the typical considered strategy. Most vendor funnel out suspect files as they move through the stack,

reducing the number of files requiring sandbox analysis. A mix of signatures, reputation, real-time emulation and heuristics enhance protection and identify advanced malware. Moreover, many security solutions has expanded deployment options with virtual and cloud-based offerings. Antiviral solutions [18] start to stop known threats then adds the next layer of defense using machine learning to detect advanced malware with both statistical analysis and behavioral analysis. Static analysis quickly compares features against those of known threats. If the file cannot be confidently convicted, it will be executed for further behavioral analysis limiting the greyware and blocking suspicious activity. Finally, convictions and IOCs are shared for enhanced protection.

### C. Anti-Antiviral Techniques

In the other hand, in order to hinder analysis, remain undetected and persist in the network, typically malware use passive and active self-defense mechanisms [17], [22], [23]:

- a- Stealth techniques: The ability to deceive any surveillance of the system by reflecting the image of a normal behavior in order to persuade the absence of any infection.
- b- Polymorphism: The ability to modify all or part of its own code to prevent any equivalent patterns.
- c- Code rewriting: The capacity to modify the code into corresponding functions.
- d- Applying encryption techniques: The procedure of masking the code to complicate the cryptanalysis .
- e- Code armouring: A number of mechanisms aiming to interrupt, delay or avoid the analysis and make the detection burdensome.
- f- Obfuscation: The fact to store codes in obscured ways to make forensics more challenging. In a  $\tau$ -obfuscation approach, the procedures remains effective for a given time or for a certain trigger.
- g- Disrupting antiviral solutions: Many techniques aiming to modify the functioning of antiviral tools and to block specific security queries.
- h- Packing: The process of compressing a software outcomes in a new altered sequence of bytes.
  - i- Anti-debugging techniques: Many techniques aiming to prevent analysts, obtaining context, attaching files and reversing code and able to detect emulation and virtual machines execution.
  - j- Steganography: The concealment of the viral payload inside another file or image.

However, these adopted techniques have many limitations and they are obviously apprehended by most antiviral solutions. First, these mechanisms usually require the combination of several techniques. Furthermore, they are difficult to implement and manage. Moreover, some of these techniques may modify the code by adding random instructions or delay the analysis but the final result is the same and the encryption procedures remain unchanged. Besides, malware are 100% of viral information in a single file. Thus, by analysis, we can speculate their operations. There are endless ways to conceive malware. At the moment, designing a truly advanced

malware, which will circumvents the known protections, is a difficult and highly competent task. In the rest of this paper, we will present a new approach of malware conception using blockchain and based on the k-ary concept.

### III. THE K-ARY MALWARE

This section presents a new category of malware denoted k-ary malware. As an alternative of holding the whole instructions constituting a malicious program in one file, this category encompasses k separate chunks which constitute a partition of the full code. Each of these programs holds only a subdivision of the instructions and reflects a regular uninfected program.

#### A. k-ary malware definition

The K-ary malware was introduced initially in 2007 [24] and has been later validated by several proof-of-concepts (POCs) [24], [25]. The formalization of this new type of malware is generalized from Cohen's model using another approach based on vector Boolean functions in order to study softwares interactions. The modalization has proved that simple and polymorphic/metamorphic infections are one way or another correspondent due to the fact that the full information is accessible after the first step of infection. Whereas, the interesting element in k-ary malware consists in the segregation of information.

Essentially, a k-ary malware is a combined virus where the viral payload is separated and distributed into k different files [24]. Each part looks like an innocent executable file and do not generate any indication of compromise (IOC). Two main categories of k-ary codes exists [17]:

- (i) Class I code: The k parts are working sequentially. Three subcategories are to be considered depending on the relation between the several parties. The execution of these k files can be dependent from the others files (*A subclass*), no part is denoting the other (*B subclass*) or semi-dependent from their execution (*C subclass*).
- (ii) Class II code: The k parts are working in parallel. Thus, all chunks have to be available and active in the system at the same period.

Furthermore, the k-ary malware are represented in Van Wijngaarden grammars defining the selection of the malware parts [26]:

$$\alpha R_m \gamma \Leftrightarrow \{\exists \omega \in (\alpha \otimes \gamma) | \omega \in C(G_m)\}$$

If the result of the selection function  $\otimes$  of two files  $\alpha$  and  $\gamma$  is a part of the code  $C$  generated by the malware  $m$  then it is a k-ary code.

#### B. Complexity

While Cohen [12] and Adleman [27] analyzed viruses with analogy to Turing machines and recursive functions and a generalized model for malicious behavior have been defined in [28] and [29], these studies do not reflect new malware interactions. For instance, the formalization of combined viruses have been well studied in [30]. Furthermore, it has

been demonstrated that the problem of detecting a k-ary malware is NP-complete [31] [32] and that the presence of all codes in memory in Class II codes and Class I (A and C subclasses) constitutes a flaw, except when using a joint rootkit technology. On the other hand, in [26], the automatic generation of K-ary codes have been detailed, sustaining their detection difficulties and their complexity. Also, in [33], k-ary codes were modulated through Join Calculus and have been demonstrated to be undecidable, except for a calculus fragment not inevitably applicable.

Therefore, given the NP-completeness of k-ary codes detection, in order to explore the feasibility of a truly undetectable malware, we will use the concept of combined viruses using the blockchain network.

#### C. Implementations

Multiple proof-of concepts confirmed the complexity of combined viruses for OpenOffice in win32 and Linux environment [32].

Moreover, the different subclasses were validated in serial ( $4 \leq k \leq 8$ ) [11] and in parallel ( $k = 4$ ) [11]. For instance, each part has been able to regenerate the missing codes under different nomenclatures.

Furthermore, a k-ary virus was implemented in Python [34] in order to share a secret key utilized to decipher the viral payload. The first use case randomly divided the key between the different parts. However, this method necessitates the availability of all parts in order to retrieve the payload. Another solution consisted in adopting Shamir's Secret Sharing with Neville/Aitken's algorithm to resolve the key and implement the k-ary virus.

### IV. THE BLOCKCHAIN POTENTIAL IN A K-ARY MALWARE

As stated previously, we will develop a k-ary malware utilizing Blockchain. This section gives an overview on blockchain networks and their features and exposes the new k-ary malware implementation and testing.

#### A. Blockchain Overview

The blockchain is a secure peer-to-peer environment used to maintain a public ledger of transactions between parties where trust is utilized to achieve consensus. This latter depends on several algorithms and typically differ according to the blockchain type and the Distributed Ledger Technology (DLT) employed. Primarily, the blockchain is an immutable data structure using blocks as memory units where each block is referenced by its hash. To characterize the transactions, the root of the Merkle tree is stored. Each block is composed by several transactions where digital signatures and cryptographic schemes are used to verify each transaction. Moreover, heterogeneous nodes are supported in the distributed network. Each node will verify and broadcast the block until reaching a consensus. The first miner to validate the blocks is rewarded. Furthermore, many types of consensus algorithms assign a penalty to misbehaving peers [35].

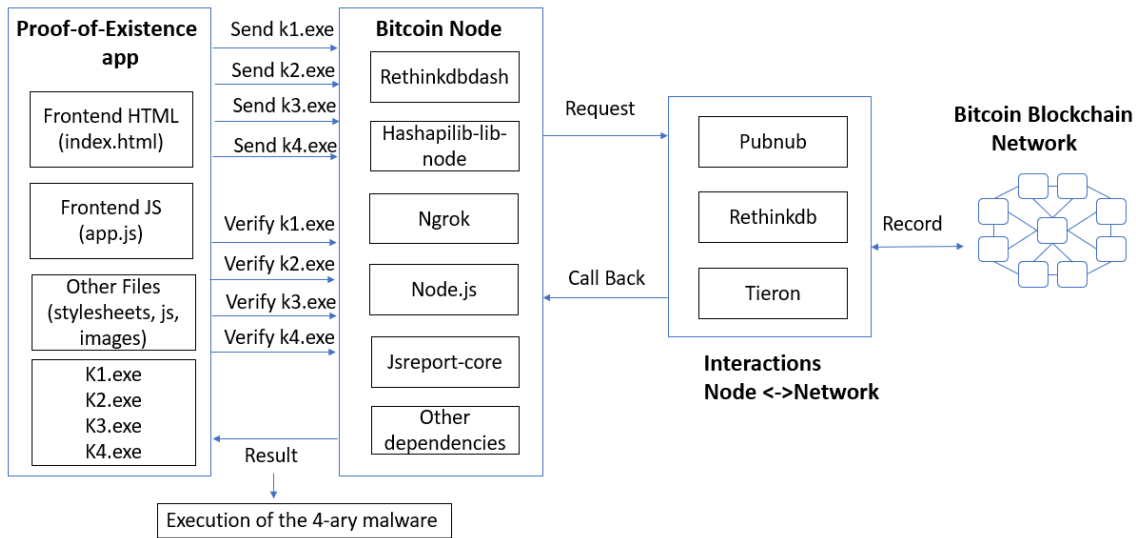


Fig. 1. 4-ary malware workflow

At the time of writing this paper, there exist more than 700 blockchain types and most of them are alternatives variants of the Bitcoin blockchain. We explored this technology by testing the three main DLTs in the market nowadays [36]–[38] namely Bitcoin, Ethereum and Hyperledger. Mainly, the difference between these networks comes from the fact that Bitcoin and Ethereum are permission-less networks whereas all parties need to be identified in the Hyperledger blockchain. In addition, the concept of smart contracts, which are function codes compiled with valid transactions, only exists in the Ethereum and Hyperledger networks.

For a long time, the blockchain technology was associated with the Bitcoin DLT based on the Proof-of-Work mechanism. However, each DLT network is characterized by its own features and consensus algorithm.

Regardless of the DLT type, the blockchain technology offers numerous security features.

Therefore, the building blocks offered a trusted platform that applications are built on top [36], [37]. Some of the blockchain's technology applications are listed below:

- **Proof-of-Existence:** Users can verify the existence of a particular content on the blockchain and that it has not been modified. Cryptographic hashes, fingerprint and a proof will be available lastingly.
- **Payment Channels:** Two parties can exchange transactions ensuring settlement and censorship resistance within a fixed deadline.
- **Crowdfunding:** Users can contribute for many causes and the incremental amount will not be spent until reaching a target.
- **Event Registration:** Users can register to an event or buy tickets through smart contracts interactions.
- **IoT:** Many IoT applications are taking advantages of blockchain networks in different use cases [39].
- **Authentication and identification:** Many companies are

```

C:\Users\Joanna>node server.js
Creating a pool connected to localhost:
Creating a pool connected to localhost:
Setting up Block Subscription...
Update subscription
Server up and listening on port
{ receiptId: "5a4f81c2b8af8a2f6d787db5", timestamp: 1515168082 }

```

Fig. 2. Node.js

leveraging DLTs for applications and entities validations [40].

The use of blockchain has been beneficial in several applications and in different fields. However, malicious entities took also advantages from this backbone. Cryptocurrencies theft and the 51% problem where a self-interested miner owns the majority of network work in a Proof-of-Work consensus (in Bitcoin and Ethereum early releases) are typical misuses. Moreover, cryptocurrencies are widely adopted by ransomware infiltrators. And in some cases, malicious contents are sold and uploaded to the blockchain encrypted and abused by the owners of the decryption key further than other mistreated scenarios and dark web applications. Furthermore, the Tor network recently leveraged the blockchain to conduct illegitimate activities [41]. Besides, in the next section, we will leverage the blockchain as a main entity to create the k-ary malware.

### B. Malware design

As we have seen previously, several attacks scenarios leveraged the blockchain to conduct fraudulent activities [42]. Whereas, for the conception of the new viral algorithms, the blockchain network is a crucial part of the new k-ary malware design.

Designing a k-ary malware [34] will lead us to a key management problem to identify each node and a key generation problem to agree on the complexity of the keys. Besides, we need to add randomization for more efficiency. To resolve these problems, we resorted to the blockchain technology.



Furthermore, payload and behavioral analysis in sandboxes are considered for advanced malware discovery. Moreover, application containment approaches are employed through agents to exclude potential offensive programs in containers and intercept their malicious activity. Finally, many agents are used for data collection and endpoints monitoring using intelligence to provide efficient protection and incident response. To fight against computer infections, several combination approaches are essentials to eliminate potential intrusions. Also, solutions integration is crucial for management, insight, activities linking and security coverage. Although these solutions are essentials, their scope of operation is limited. Furthermore, the usefulness of the behavioral detection and the will to obtain a lower probability as well as the compromises and algorithmic choices may completely inhibit the essential property of the detection. The fundamental lever is the human factor and the security policies in place. According to the theory of computability, some problems are not calculable and the problem of viral detection is one of the undecidable problems.

In this paper, we utilized the Blockchain in order to explore the feasibility of a new undetectable malware. We based our malware on k-ary codes which have been demonstrated to be NP-complete. We have developed a 4-ary malware and tested it in real time where each chunk of the code interacts with the Bitcoin network to be validated and to make sure that it belongs to our malicious software. Therefore, the blockchain network provided an elegant solution to retrieve the multiple parts of the malware, making sure of their authenticity and integrity without worrying about the generation, the management and the storage of the keys.

The next step consists in leveraging smart contracts functions to enhance our k-ary malware and add more complexity. Besides, we will tackle its formalization and validate it in real time against advanced and sophisticated endpoints protections.

## REFERENCES

- [1] K. Zetter, "Nov. 10, 1983: Computer 'Virus' Is Born," <https://www.wired.com/2009/11/1110fred-cohen-first-computer-virus/>, 2009, [Online; accessed 11-November-2017].
- [2] T. M. Chen and J.-M. Robert, "The evolution of viruses and worms," *Statistical methods in computer security*, vol. 1, 2004.
- [3] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *IEEE Security & Privacy*, vol. 9, no. 3, pp. 49–51, 2011.
- [4] J. P. Farwell and R. Rohozinski, "Stuxnet and the future of cyber war," *Survival*, vol. 53, no. 1, pp. 23–40, 2011.
- [5] R. Abreu, F. David, and L. Segura, "E-banking services: Why fraud is important?" 2016.
- [6] A. Peterson, "The sony pictures hack, explained," *The Washington Post*, vol. 1, 2014.
- [7] L. Constantin, "New locky ransomware version can operate in offline mode," 2016.
- [8] R. Dobbins, "Mirai iot botnet description and ddos attack mitigation," *Arbor Threat Intelligence*, vol. 28, 2016.
- [9] A. Greenberg, "The wannacry ransomware hackers made some real amateur mistakes," 2017.
- [10] R. Richardson and M. North, "Ransomware: Evolution, mitigation and prevention," *International Management Review*, vol. 13, no. 1, p. 10, 2017.
- [11] É. Filiol, *Techniques virales avancées*. Springer, 2007.
- [12] F. Cohen, "Computer viruses: theory and experiments," *Computers & security*, vol. 6, no. 1, pp. 22–35, 1987.
- [13] G. Hoglund and J. Butler, *Rootkits: subverting the Windows kernel*. Addison-Wesley Professional, 2006.
- [14] M. E. Russinovich, D. A. Solomon, and A. Ionescu, *Windows internals*. Pearson Education, 2012.
- [15] J. Moubarak, M. Chamoun, and E. Filiol, "Comparative study of recent mea malware phylogeny," in *Computer and Communication Systems (ICCCS), 2017 2nd International Conference on*. IEEE, 2017, pp. 16–20.
- [16] J. Moubarak, E. Filiol, and M. Chamoun, "Comparative analysis of blockchain technologies and tor network: Two faces of the same reality?" in *CSnet , 1st Cyber Security in Networking Conference*. IEEE, 2017.
- [17] E. Filiol, *Computer viruses: from theory to applications*, 2006.
- [18] McAfee, "Powerful advanced threat detection," 2015.
- [19] PaloAlto, "Next Generation Firewall," 2017.
- [20] FireEye, "Endpoint," 2017.
- [21] Kaspersky, "Endpoint Security for Business," 2017.
- [22] E. Filiol, "Malicious cryptology and mathematics," in *Cryptography and Security in Computing*. Intech, 2012.
- [23] I. You and K. Yim, "Malware obfuscation techniques: A brief survey," in *Broadband, Wireless Computing, Communication and Applications (BWCCA), 2010 International Conference on*. IEEE, 2010, pp. 297–300.
- [24] E. Filiol, "Formalisation and implementation aspects of k-ary (malicious) codes," *Journal in Computer Virology*, vol. 3, no. 2, pp. 75–86, 2007.
- [25] M. Dalla Preda and C. Di Giusto, "Hunting distributed malware with the  $\kappa$ -calculus," in *Fundamentals of Computation Theory*. Springer, 2011, pp. 102–113.
- [26] G. Gueguen, "Van wijngaarden grammars, metamorphism and k-ary malwares," *arXiv preprint arXiv:1009.4012*, 2010.
- [27] L. M. Adleman, "An abstract theory of computer viruses," *Advances in Crypto*, 1998.
- [28] Z. Zuo and M. Zhou, "Some further theoretical results about computer viruses," *The computer journal*, vol. 47, no. 6, pp. 627–633, 2004.
- [29] G. Bonfante, M. Kaczmarek, and J.-Y. Marion, "On abstract computer virology from a recursion theoretic perspective," *Journal in computer virology*, vol. 1, no. 3, pp. 45–54, 2006.
- [30] E. Filiol, "Metamorphism, formal grammars and undecidable code mutation," *International Journal of Computer Science*, vol. 2, no. 1, pp. 70–75, 2007.
- [31] Filiol, "Malware of the future," 2015.
- [32] D. de Drézigué, J.-P. Fizaine, and N. Hansma, "In-depth analysis of the viral threats with openoffice.org documents," *Journal in Computer Virology*, vol. 2, no. 3, pp. 187–210, 2006.
- [33] G. Jacob, E. Filiol, and H. Debar, "Formalization of viruses and malware through process algebras," in *Availability, Reliability, and Security, 2010. ARES'10 International Conference on*. IEEE, 2010, pp. 597–602.
- [34] A. Desnos, "Implementation of k-ary viruses in python," *Hack lu*, 2009.
- [35] G. O. Karame and E. Androulaki, *Bitcoin and Blockchain Security*. Artech House, 2016.
- [36] A. M. Antonopoulos, *Mastering Bitcoin: unlocking digital cryptocurrencies*. O'Reilly Media, Inc., 2014.
- [37] A. Bahga and V. Madiseti, "Blockchain applications: A hands-on approach," 2017.
- [38] I. Cloud, "Blockchain," <https://console.bluemix.net/catalog/services/blockchain/>, 2017.
- [39] Postcapes, "Blockchains and the Internet of Things," 2017.
- [40] LTP, "22 companies leveraging blockchain for identity management and authentication,"
- [41] BlockchainBlog, "Blockchains and the Internet of Things," <https://blog.blockchain.com/tag/tor/>, 2017, [Online; accessed 12-November-2017].
- [42] H. Patel, "Blockchain-ware: Next stage of malware evolution," 2017.
- [43] J. Riordan and B. Schneider, "Environmental key generation towards clueless agents," *Mobile agents and security*, vol. 1419, pp. 15–24, 1998.
- [44] Pubnub, "Build a Proof of Existence Service in the Blockchain," 2017.