ANALYSIS

BLOCKCHAIN SECURE CLOUD: A NEW GENERATION INTEGRATED CLOUD AND BLOCKCHAIN PLATFORMS – GENERAL CONCEPTS AND CHALLENGES

JOANNA KOŁODZIEJ

is the Professor of computer science and Head of the Computer Science Department at Cracow University of Technology. She works also in Research and Academic Computer Network (NASK) Institute. Prof. Kołodziej serves as the President of the Polish Chapter of IEEE Computational Intelligence Society. She participated in several international and national projects including ECONET, 7FP and PARAPHRASE 7FP Grants. Currently, she is the Principal Investigator of Horizon 2020 cHiPSet Cost project IC1406 (chipset-cost.eu). She is the leader of the Polish consortium of BalticSatApps EU InterReg project (http://balticsatapps.eu/). Prof. Kołodziej is the author and editor of 20+ books and the author of 170+scientific publications in top world international journals in the area of computer science and applied mathematics.

ANDRZEJ WILCZYŃSKI

is an Assistant Professor at Cracow University of Technology and Ph.D. student at AGH University of Science and Technology. The topics of his research include blockchain-based modelling in distributed computing, cloud computing, in particular data and resource virtualization, tasks scheduling in cloud computing and broadly defined security in these areas.

DAMIÁN FERNÁNDEZ-CERERO

received the B.E. degree and the M.Tech. degrees in Software Engineering from the University of Sevilla. In 2014, he joined the Department of Computer Languages and Systems, University of Seville, as a Ph.D. student. In 2016 he was invited by at ENS-Lyon and in 2017 at Cracow University of Technology to work in saving energy solutions for cloud infrastructures. Currently he both teaches and conducts research at University of Sevilla. He has worked on several research projects supported by the Spanish government and the European Union. His research interests include energy efficiency and resource scheduling.

ALEJANDRO FERNÁNDEZ-MONTES

received the B.E. degree, M. Tech. and International Ph.D. degrees in Software Engineering from the University of Sevilla, Spain. In 2006, he joined the Department of Computer Languages and Systems, University of Sevilla, and in 2013 became Assistant Professor. In 2008 and 2009 he was invited to the ENS-Lyon, in 2012 to the Universitat Politecnica de Barcelona and in 2016 to Shanghai Jiao Tong University for share experiences and knowledge in saving energy solutions for Data Centers. His research interests include energy efficiency in distributed computing, applying prediction models to balance load and applying on-off policies to Data Centers.

Introduction

As we delve deeper into the 'Digital Age', we witness an explosive growth in the volume, velocity, and variety of data available on the Internet. For example, in 2012, about 2.5 quintillion bytes of data were created each day. The data originated from multiple types of sources, including mobile devices, sensors, individual archives, social networks, the Internet of Things, enterprises, cameras, software logs, etc. Such 'data explosions' have raised one of the most challenging research questions of the current Information and Communication Technology (ICT) era: how to effectively and optimally manage such large amounts of data and identify new ways to analyse them in order to unlock information.

Millions of financial transactions realised each day in today's global market generate hundreds of petabytes of sensitive heterogeneous data, which requires it to be processed efficiently (stored, distributed, and indexed) in a way that does not compromise end-users' Quality of Service (QoS) in terms of data availability, data privacy, data search delay, data analysis delay, and the like. Many of the existing ICT systems that store, process, distribute, and index hundreds of petabytes of heterogeneous data fall short of this challenge or simply do not exist yet. There has been a paradigm shift in executing high-performance large data applications from physical hardware- and software-enabled platforms managed locally, which should be processed, analysed and stored in safe ICT environments.

The main research challenge in the ICT support of the financial markets is the development of a next generation financial technology for a secure use of electronic currencies and a secure network technology for system user communication, as well as data processing and storage without the involvement of third parties. To deal with the security aspects of financial virtual transactions, *blockchain* technology has been proposed. Blockchain can be defined as a public ledger network for secure online transactions with virtual currencies. Transaction records are encrypted by using cryptographic methods and executed in a distributed computer network as blockchain software.

The blockchain model has been gaining popularity since 2008, when the first electronic money protected through the cryptographic mechanisms (cryptocurrencies) was introduced. The first cryptocurrency to use a blockchain-based approach was Bitcoin (Il-Kwon et al., 2014). These currency blockchain systems store the value attached to a digital wallet—an electronic device (or software) that allows realising electronic transactions.

Blockchain transactions are finalized through an authentication process, where the customer who borrows virtual money creates a block of transactions. This block is periodically updated and reflected in the electronic money transaction details to share the latest transaction detail block (Armknecht et al., 2015).

Blockchain can be successfully utilised in diverse areas, including the financial sector and the ICT computational environment, such as computational clouds (Christidis and Michael, 2016), (Huh et al., 2017). Cloud computing gives application developers the ability to marshal virtually infinite resources with an option to pay-per-use and as needed and does not require upfront investments in resources that may never be optimally used. Once applications are hosted on cloud resources, users are able to access them from anywhere, at any time, using devices ranging from mobile devices (smartphones, tablets) to desktop computers. The data centre cloud provides virtual centralisation of applications, computing, and data. While cloud computing optimises the use of resources, it does not (yet) provide an effective solution for the secure hosting of large data applications (Singh et al., 2016).

In this paper, we define the generic model and the main characteristics of the blockchain network. We present it as a reference infrastructure, which can be easily combined with other large-scale distributed computational environments. We briefly discuss the concept of integration of blockchain with cloud platforms in order to improve the security of data storage as well as resource, data and user management in both environments.

Blockchain origins

Blockchains can be defined as distributed digital ledgers of cryptographically signed transactions that are grouped into blocks. Each block is cryptographically linked to the previous one after validation and undergoing a consensus decision." (Yaga et al., 2018) Thanks to encryption technologies, the single point of failure caused by an authorised third party has been overcome when it comes to verifying the authenticity of transactions.

The blockchain model leverages many features of the 'Peer2Peer' (P2P) model. This broker-free approach enables users to not incur avoidable costs related to third-party centralised authorisation operations. In this model, security standards are higher and transactions are committed faster as they are automatically accepted and saved by multiple agents. It makes it harder for hackers to exploit vulnerabilities of the system, thus reducing costs of security-related tasks. Furthermore, transactions can be easily made public and open access.

Figure 1 shows the basic components of the blockchain P2P architecture. There are many variations of this basic conceptual design, including other features, but the diagram is a useful way to describe the way blockchains work.

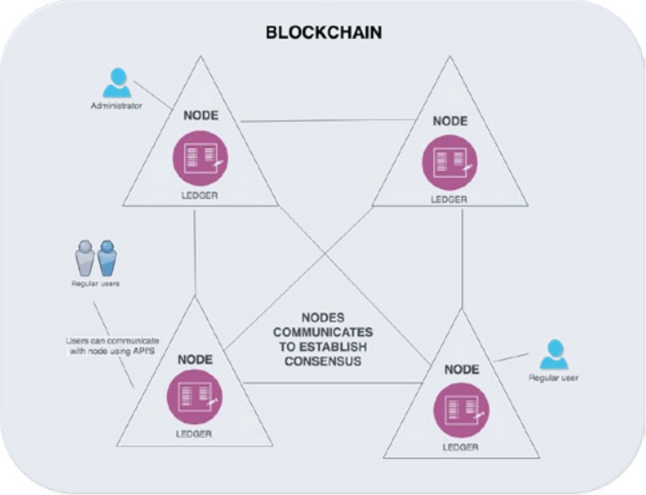


Figure 1. Blockchain distributed architecture

Hashes

Cryptographic hash functions are the main component of the blockchain model. These functions are widely utilised for several use cases, e.g. encrypting the data present in a block. Almost any input of any size (e.g., a picture or a text file) can be processed with the use of 'hashing'. The main goal of the hashing method is to compute a 'unique' static-sized output, i.e., a 'message digest'. Every single change in the input files may produce a totally different output message digest. Furthermore, two inputs cannot result in the same output (computationally), which means that hash algorithms are 'collision resistant' (second pre-image resistant).

The fast-computing SHA–256 algorithm (Secure Hash Algorithm with an output size of 256 bits) is a well-known and widely used algorithm supported by a majority of computing nodes and utilised by a lot of blockchain-based models. The NIST Secure Hashing website (Dang, 2014) contains FIPS specifications for all NIST-approved hashing algorithms. One of them, the SHA–256 algorithm, is specified in the Federal Information Processing Standard (FIPS) 180–4 (Dang, 2014).

Ledgers

Ledgers are composed of a set of transactions. Each *node* has a local copy of this set of transactions, i.e., the ledger. By the same token, a blockchain is usually composed of a set of *nodes*. The exchange of goods and services has been stored historically in analogue (pen-and-paper) ledgers. With the adoption of new computing paradigms, these analogue ledgers have been substituted with records in large centralised databases. These records are generated by a collection of users who entrust the operation of such databases to 'trusted' external agents, which actually own the data and ledgers. However, this centralized ledger approach has some disadvantages which include:

- The centralised agent is a single point of failure of the whole system. This means that at least the owner needs a backup system (or user) in case of loss or destruction.
- Each committed transaction should be validated by the central third-party agent. This means the validity of the transactions are only backed by the owner whom all users must trust.

- In the same way, all users must trust that this central agent corroborates the completeness of the ledger, since some transactions may be lost (purposefully or due to failures upon reception).

However, it should be noted that a vast majority of third-party 'trusted' agents and companies do backup transactions, which is in their best interest and the interest of their final users, validate the committed data, including all valid transactions.

Blocks

Each of the nodes in the blockchain may receive candidate transactions submitted by end-users. These transactions are then propagated to other nodes in the working group network. This operation, however, does not actually save the transaction in the blockchain. Subsequent to this process, mining nodes need to add the aforementioned transactions to the blockchain. Until then the committed transactions wait in the 'transaction pool' (a queue).

As mentioned before, the mining nodes are responsible for keeping the blockchain up-to-date by publishing freshly committed blocks. This process performs the actual operation of adding transactions to the blockchain. Thus, a 'block' is composed of validated transactions. To this end, the providers of transactions, who are shown in the input values of each transaction, must cryptographically sign the transaction to ensure its 'legitimacy', meaning that each of them had access to the appropriate private key. No blocks containing invalid transactions will be accepted in the blockchain. To this aim, the rest of the mining nodes in the network check the validity of each and every transaction in the published block. Once a block is created, it must be hashed. To this purpose, a 518 digest, which represents the block, will be created. The immutability of data is ensured by this method since even a change in a single bit of the block would drastically change the generated hash. In addition, a copy of the hash of every block is shared among all the nodes in order to improve security. This system prevents any change since every node can check if the hash matches.

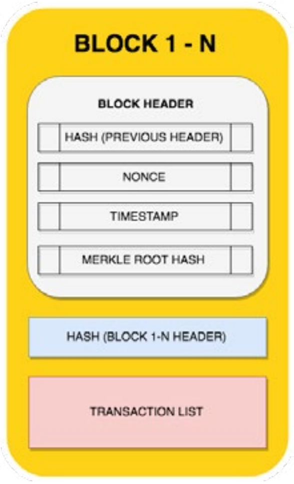


Figure 2. Block generic model

Each block typically consists of the following components:

- The block number, also known as 'block height'
- The current block hash value
- The previous block hash value
- The Merkle tree root hash
- A timestamp
- The size of the block
- A list of transactions within the block

The generated hash is stored in a data structure called 'Merkle tree' instead of the header of the block. The hash values of the gathered data are combined by the Merkle tree until there is a singular root, called 'Merkle tree root hash'. The presence of transactions within blocks and their summary can be efficiently verified by means of the aforementioned root. In addition, this data structure enables the system to detect any changes to the underlying data, therefore assuring that the data sent through the network is valid. Figure 3 shows an example of a Merkle tree:

- The bottom row shows the transaction data, i.e., the data to be summarised in the blockchain.
- The second row from the bottom represents the hashing process.
- The resulting hashed data is then combined and hashed. This is shown in the third row starting from the bottom.
- The top row represents the root hash, which hashes and combines H4 with H5. The root hash is created from the set of hashes containing all previous combinations and hashes.

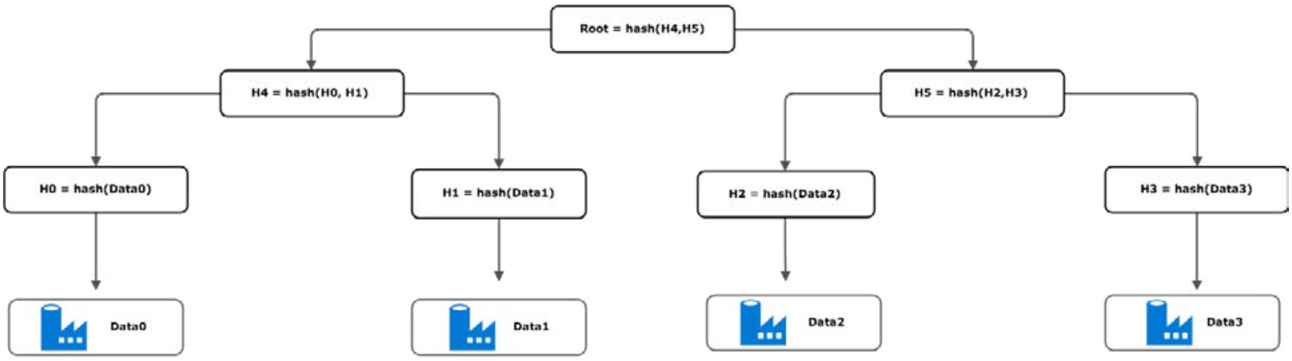


Figure 3. Merkle tree

Blockchain processes

In most blockchain platforms, the nodes of the blockchain network are owned by different organisations. The nodes may communicate with each other to agree on ledger content, and no central authority is required for both coordination and validation of all transactions.

Several algorithms have been proposed to solve the problem of reaching the agreement between nodes, i.e., the *consensus*. The blockchain receives transaction requests, which are submitted by users, to perform the operation it has been designed for. As a result of the execution of such a transaction, one or more ledgers store a record of the transaction which will never be modified or deleted. With this process, the *immutability* of the blockchain is achieved.

Blockchain security

Blockchain platforms are network environments where transaction data and parameters (value, state) are close to business logic. Blockchain transactions are mainly based on cryptographic and other mathematical models implemented for trading partners. The most popular cryptographic methodology used for blockchain transactions and data is asymmetric-key cryptography (Stallings, 1990) (also referred to as public/private-key cryptography). In this model, there is a pair of keys –a public key and a private key – used for signing the transactions and verifying the signatures in the following way (Bozic et al., 2016):

- Private key is used to generate transaction digital signatures.
- Public key is used to verify the signature generated with private key.

The public key may be made known to many users without affecting the security of the whole transaction. The private key, on the other hand, must be made known just to the key owner. Asymmetric cryptography guarantees that private key cannot be determined based on the knowledge of the public key.

Integration of blockchain with cloud environments

Cloud computing (Wang et al.) assembles large networks of virtualized services: hardware resources (CPU, storage, and network) and software resources (databases, message-queuing systems, monitoring systems, load-balancers). In the industry, these services are referred to as 'Infrastructure as a Service' (IaaS), 'Platform as a Service (PaaS)', and 'Software as a Service' (SaaS). Cloud computing services are hosted in large data centres, often referred to as 'data farms'.

Based on resource and data management and the related security and privacy issue, we can distinguish three main types of cloud platforms: (i) public cloud, (ii) private cloud, and (ii) hybrid cloud. Public clouds offer unlimited access to shared data and resources for a wide group of users, but there is no guarantee that users' data will be protected. Access to resources and data in private clouds is restricted and each user must be validated through strong authorisation and authentication procedures. Private cloud clusters are usually owned by enterprises and work under specific cloud standards. Hybrid clouds seem to be an ideal model of integration of the many private clouds into a joint global infrastructure. Such integration is done through the upper-level public layer. The main problem with that model is to reach an agreement among private cloud providers to work under a unified public cloud standard. Therefore, the 'many cloud model', where the distributed private cloud clusters are connected by using the standard P2P network (see Figure 4), is a much more realistic scenario.

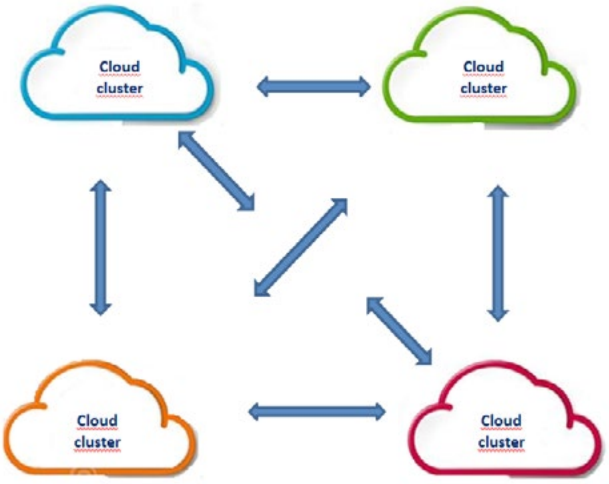


Figure 4. P2P-based 'Many clouds' architecture

It can be observed that a similar model works for the blockchain network, which was the first reason for trying to integrate both environments in order to improve the security policies in global clouds.

There are two main methods of integration of the cloud with blockchain platforms:

1. Using cloud for the development of blockchain applications and supporting the integration with enterprise networks (private clouds) to facilitate storage, replication and access to transactional data;
2. Using blockchain methods to improve the security of task, user and data management in the clouds.

Challenges, special conditions and main problems related to data and users' privacy, along with the recent ideas and developments are briefly discussed in the section below.

Cloud support for blockchain transactions and data – challenges and requirements

The number of transactions in blockchain networks can be enormous. The large volumes of generated data need scalable data processing services. Elasticity and scalability are some of the most important functionalities of the cloud systems to provide on-demand cloud resources for dynamically changing workload.

Public clouds can offer a large-scale network of resources available for the customers who pay only for the utilised ones. Private clouds usually need to be optimised for handling large data sets.

From the security perspective, cloud systems can effectively hide the physical location of data. Tuning activities can be carried out continuously with minimal impact on deployed applications, which is crucial for an efficient implementation of most of the blockchain algorithms. Any blockchain system must take into account data sovereignty rules and store and process data only in the locations permitted by the regulations. It means that the cloud service provider allows their customers to have control over the locations in which their data is stored and processed.

Another important issue regarding blockchain networks is system resilience and fault tolerance. It means that a failure of any single node in the blockchain network should not affect the work of the whole system. Cloud services help in such cases through the replication of data stored in data centres and the use of multiple software applications.

Finally, the implementation of blockchain algorithms in clouds may improve the security of the blockchain system itself. Software can be centrally maintained in a distributed cloud environment with data stored on a local data server. The recent examples of such successful integration of the blockchain with cloud platforms are Oracle Blockchain Cloud Service project (Oracle, 2017) and iEx.ec project (iEx.ec, 2018).

Blockchain support for the cloud users, task and data management – new ideas

The most recent technological developments and anonymization of the user's information and data in the cloud environments are inspired by blockchain technologies (Bozic et al., 2016). Blockchain seems to be a promising methodology for ensuring anonymity in large-scale clouds, an electronic wallet for user anonymity (Park and Park, 2017). Such an electronic wallet is installed when using blockchain technology, and after that it must be securely deleted from the system to avoid the private user's information being accessed by third parties.

Another new idea is to use blockchain proof of concept algorithms for secure data and task scheduling in the cloud. For example, the 'many clouds' model is used for illustrating the distributed P2P cloud cluster architecture. Each node in that P2P network corresponds to the cloud Service Provider (SP). The SP node may have a complex internal architecture: one SP node may be the master node for the local data and computational servers (slave nodes). The execution of the generated optimal schedule can be additionally monitored by the blockchain system in order to generate the recommendation list of the data storage servers, cloud services and cloud resource providers. This is totally new concept, which is currently one of our research tasks in our cloud development work.

Conclusions

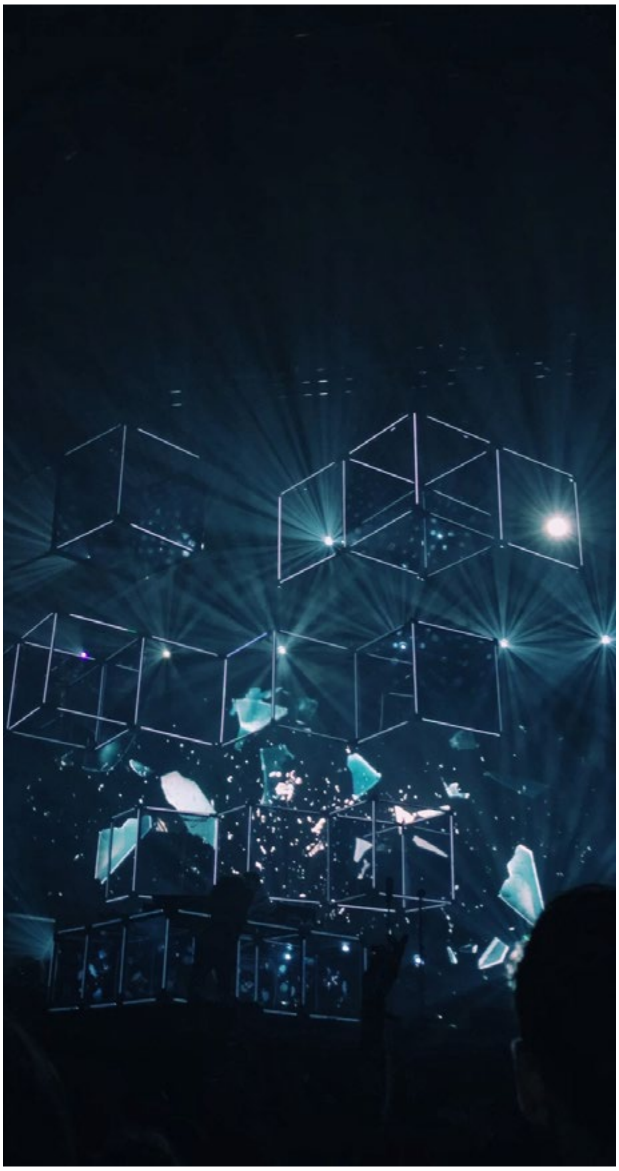
Blockchain is a popular financial technology, which uses ICT environments for virtual financial transactions using cryptocurrencies (e.g. Bitcoin). Blockchain customers store their transaction records in the blockchain P2P network, which effectively utilises the computing resources of its peers. A proof of work and a proof of stake are blockchain consensus algorithms that are used to improve the security of blockchain transactions.

In this paper, we briefly discussed the benefits of integrating the blockchain network with the elastic, scalable cloud environment in order to enhance the trustfulness of data servers and the security of data and user management. We also identified the challenges posed by this integration process.

The new concept that we propose is to use blockchain algorithms for monitoring the execution of the security-aware task scheduling in the cloud, which is one of the most important research topic in today's cloud and fog computing. We believe that the new blockchain-based scheduling model will allow us to overcome the problem related to the implementation of the existing models in the real-life scenarios (Kołodziej et al., 2014).

Acknowledgement

This paper is based upon work done as part of COST Action IC1406 High-Performance Modelling and Simulation for Big Data Applications (cHiPSet), supported by COST (European Cooperation in Science and Technology). ■



REFERENCES

Il-Kwon, L. Young-Hyuk, K. Jae-Gwang, L. and Jae-Pil, L. (2014, June 30-July 3). The Analysis and Countermeasures on Security Breach of Bitcoin. *Proceedings of the International Conference on Computational Science and Its Applications.* Guimarães, Portugal. Springer International Publishing: Cham, Switzerland.

Christidis, K. and Michael, D. (2016). Blockchains and Smart Contracts for the Internet of Things. *IEEE Access, 4.* pp. 2292–2303.

Huh, S. Sangrae, C. and Soohyung, K. (2017, February). Managing IoT devices using blockchain platform. *Proceedings of the 2017 19th International Conference on Advanced Communication Technology (ICACT).* Bongpyeong, Korea. pp. 19–22.

Armknecht, F. Karame, G. Mandal, A. Youssef, F. and Zenner, E. (2015). Ripple: Overview and Outlook. In Trust and Trustworthy Computing. *Conti, M., Schunter, M., Askoxylakis, I., Eds.* Springer International Publishing, Cham, Switzerland. pp. 163–180.

Singh, S. Jeong, Y.-S. Park, J.H. (2016). A survey on cloud computing security: Issues, threats, and solutions. *J. Netw. Comput. Appl.* 75. pp. 200–222.

Dang Q. H. (2014). Secure Hash Standard. *Federal Inf. Process. Stds. (NIST FIPS) - 180-4.* Retrieved from https://www.nist.gov/publications/secure-hash-standard

CCAB. (2017). *Cloud Customer Architecture for Blockchain.* Cloud Standards Customer Council, http://www.cloud-council.org/deliverables/cloud-customer-architecture-for-blockchain.htm

Zeng, Z. Dai, H.-N. Xie, S. and Wang, H. (2017). An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. *Proc. Of the 2017 IEEE 6th International Congress on Big Data.* pp. 557-564.

Yaga, D. Mell, P. Roby, N. and Scarfone K. (2018). Blockchain Technology Overview. Draft NISTIR 8202. Retrieved from https://csrc.nist.gov/CSRC/media/Publications/nistir/8202/draft/documents/nistir8202-draft.pdf

Park J.H. and Park J.H. (2017). Blockchain Security in Cloud Computing: Use Cases, Challenges, and Solutions. *Symmetry.* 9(164). Retrieved from www.mdpi.com/2073-8994/9/8/164

Bozic, N. Guy, P. and Stefano, S. (2016). A tutorial on blockchain and applications to secure network control-planes. SCNS IEEE.

Oracle (2017). Oracle Blockchain Cloud Service project. Retrieved from https://cloud.oracle.com/blockchain

iEx.ec (2018). Retrieved from https://iex.ec/

Stallings, W. (1990). *Cryptography and Network Security: Principles and Practice.* Prentice Hall. p. 165.

EU DP (2018). 2018 reform of EU data protection rules, Retrieved from https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en

Wang, L. Ranjan, R. Chen, J. and Benatallah B. (eds.) (2011). *Cloud computing: methodology, system, and application.*, CRC Press. Taylor & Francis.

Kolodziej, J. Khan, S.U. Wang, L, Kisiel-Dorohinicki, M. Madani, S.A, Niewiadomska-Szynkiewicz, E. Zomaya, A.Y. and Xu, C.-Z. (2014). *Security, energy, and performance-aware resource allocation mechanisms for computational grids.* Future Generation Comp. Syst. 31: 77-92