"DECENTRALIZED INTERNET PRIVACY: TOWARDS A BLOCKCHAIN FRAMEWORK FOR HEALTHCARE"

Research-in-Progress

Track N° 37

Karagiannis, Stelios, Ionian University, Corfu, GR, c15kara@ionio.gr

Magkos, Emmanouil, Ionian University, Corfu, GR, emagos@ionio.gr

Abstract

This research concerns the new environment of computer networks and issues of privacy. Modern technologies and systems handle a large amount of personal information. Personal information and sensitive data are usually handled by third parties. Attacks and data breaches on large information systems, engage the danger of privacy breaches and violations. User profiling and data extraction using modern network flow methods reveal privacy issues and violations. Privacy terms as stated, are not sufficient in protecting users from data breaches and privacy violations, thus rendering more urgent the need for users to be constantly informed of any information disclosures, regaining in this way the control of their personal information.

Decentralized models for privacy, apart from their benefits, might reveal new privacy issues and specific approaches have to be analysed. Blockchains alongside other challenges, might improve privacy properties, but could also lead to the opposite result, since the shared ledger will be always available. On the other hand, TTP-based approaches have a single point of failure and often suffer from data breaches. Moreover, blockchain approaches until now are not adequately evaluated in terms of privacy. In this paper, a model is proposed for providing sensitive information in healthcare environments using blockchains, in order to give the complete control of personal data to the users. During the disclosures, it is important for the patient to give consent and to have the adequate notifications. From this perspective, blockchain technology enhances transparency, significantly in keeping users informed about any disclosures.

Keywords: Privacy, Blockchains, Decentralized models, Surveillance

1 Introduction

Our world highly depends on third parties for keeping information and personal data secure as pointed from Modi and Patil (2016). Trusted third parties (TTPs) have the obligation of keeping information safe from breaches, respecting the privacy policies. However, data breaches are constantly conducted, focused mostly on gathering information during these breaches and selling personal information illegally. Vulnerabilities in TTP-approaches still exist, since even if distributed, are still vulnerable to DDoS attacks. Moreover, third parties usually enhance actions of surveillance according to specific laws that might apply. Nowadays, the usage of internet services and levels of centralization are higher than before. Individuals keep most of their personal data in specific trusted third parties, which usually monitor traffic. Furthermore, Privacy terms created by third parties can be complex or confusing, thus difficult to be fully understood for the average user. Common users have little or no control over their personal information and data.

New technologies like distributed ledger technology, also referred to as blockchain technology can bypass the third-party trust parameter (Jacobovitz, 2016), ensuring the validation of proof of work and validating the transactions. Blockchains are widely used in cryptocurrencies like Bitcoin, but it is under research how blockchains can help in other transactions such as data indexing, or whether it is possible to use blockchains as framework or platform for applications. As suggested by Zyskind and Nathan (2015), blockchain technology might guarantees privacy, while giving more of personal assets control to the users. Decentralized applications and networks might be a solution as to building privacy-preserving systems and avoiding censorship.

Particularly in healthcare, where personal information is sensitive, the disclosures must be handled carefully, while preserving the required privacy properties and have methods for providing the adequate consent. Research in blockchains and decentralized models is ongoing with significant work by Zyskind, Nathan (2015) and Lee (2010). Specifically, Lee (2010) focus in how blockchains can create a distributed web, applications, services and frameworks using blockchain technology. Significant blockchain implementations other than Bitcoin are Ethereum (Buterin, 2014), Maidsafe (Paul et al., 2014), Decent (Jahid et al., 2012), including Rethink and IPFS among others. However, many of the services have not been supported by academic research, while most of the projects only publish white papers. This confirms that blockchain and decentralized applications are in a very early stage.

1.1 Main Contribution and Novelty

Setting up privacy requirements in blockchains, as well as considering the privacy properties and the benefits that blockchains might offer is one of the primary areas that this ongoing research aspires to investigate. In addition to that, blockchains apart from privacy issues, also reveal issues related to performance, scaling and high consumption of resources. Research in this field might lead to significant findings as to the development of decentralized applications in large-scale networks and big databases. Furthermore, decentralized models and their approaches for bypassing third party trust are assessed for handling the control of personal data by their possessors. More specifically, In this specific research:

1. We show how specific privacy and security needs of particular user groups might influence the design of a decentralized privacy-preserving platform
2. It is described how applications can be developed using blockchains in privacy-preserving systems where maintaining privacy is important, as in the case of healthcare services.
3. We propose a blockchain-based framework focusing in bypassing conducted privacy breaches along with ongoing data breaches. In addition we compare the framework with the common network model.

4. The importance of decentralized models is examined, highlighting privacy issues of the future web structure. In these decentralized models, networks can be trustless, in contrast to TTP-based approaches which may ultimately have a negative impact on users' privacy needs and requirements.

For the evaluation of the proposed blockchain-approach, a decentralized application will be developed, considering the specific environment of a decent healthcare environment, extracting quantitative results according to privacy and security. An open platform for creating blockchain applications will be used, called *Multichain*. *Ethereum* will be also considered as a possible solution for developing the decentralized application.

2 Privacy and Specific User needs: The case for Healthcare

Privacy policies are implemented bearing in mind the privacy rights, user demands and the specific regulations that apply. Privacy depends on attributes, needs and requirements of each different user group. Even if there are several frameworks for describing the privacy demands of each group, specific terms and regulations apply in different social environments. A large part of this research focuses on how health care data is managed since the requirement in privacy is critical in handling sensitive information in healthcare systems. Handling personal data to a third party is usually optional for the system usability, yet many times information disclosure is mandatory.

Quantifying privacy requires a fuller understanding of fundamental privacy notions (Pfitzmann and Köhntopp, 2017), including the attributes of anonymity, unlinkability, undetectability and observability. While *anonymity* is about hiding the identity from the entity which executed a specific action, *unlinkability* ensures that while sending specific data the corresponded recipient will not be linked by others. *Unlinkability* ensures that while sending specific data the corresponded recipient will not be linked by others. In addition, *unobservability* ensures that a user might use resources bypassing the observation from others and especially from third parties or tracked while using a service. As modern communications evolve and grow more complex, ensuring such properties becomes increasingly difficult. Each privacy property describes a different option for data minimization (Pfitzmann and Hansen, 2010). *Privacy enhancing technologies (PETS)* usually aim at safeguarding privacy properties. Conventional analysis is under research focusing mostly on properties of anonymity and unlinkability.

In order to ensure privacy, users must have an efficient way of controlling whether to conceal personal data from unauthorized parties (Buttyan and Hubaux, 2007). The sort of information that is hidden determines which of the privacy properties may apply. Setting up a basic framework and specifying the privacy requirements of each group is mandatory. Extracting quantitative indicators for measuring the ability of blockchain technology in protecting privacy is something complex considering the underlined internal personalized connections of the users. The final result is a mix of connected entities and objects within a complex network of interactions.

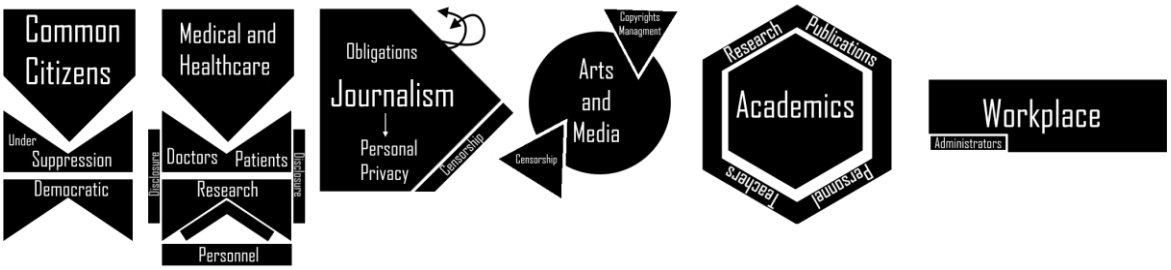

Figure 1. User groups

User groups are presented in Figure 1 according to the specific similar needs these user groups have. The user profiles that apply in each user g roup have different privacy requirements. More spe-

cifically, the basic privacy properties determine the data minimization which every user group requires. In reality, everyone uses the network according to personal unique needs, implying that every user has different demands in terms of privacy. For adequately succeed in meeting these requirements, specific privacy properties apply. Describing and setting up a framework that will abide by each user group's requirements in privacy, is a step for describing the level of utilization that each user needs along with the specific privacy properties which have to be applied. For having sufficient quantitative results about privacy in the proposed approach, specifying the privacy requirements of the specific user group is mandatory.

The proposed use case is focused in healthcare environments. Healthcare systems usually automate the transactions between organizations such as sharing medical records, images and personal information. *Electronic Healthcare Systems (EHR)* are information systems responsible for holding medical records and necessary transactions as seen in Figure 2. Personalized medical and health treatments are possible using this vast amount of data. The main challenges that occur in these systems involve data security and privacy, while maintaining a sufficient level of performance, scalability and low system resources usage (Agrawal et al., 2002).

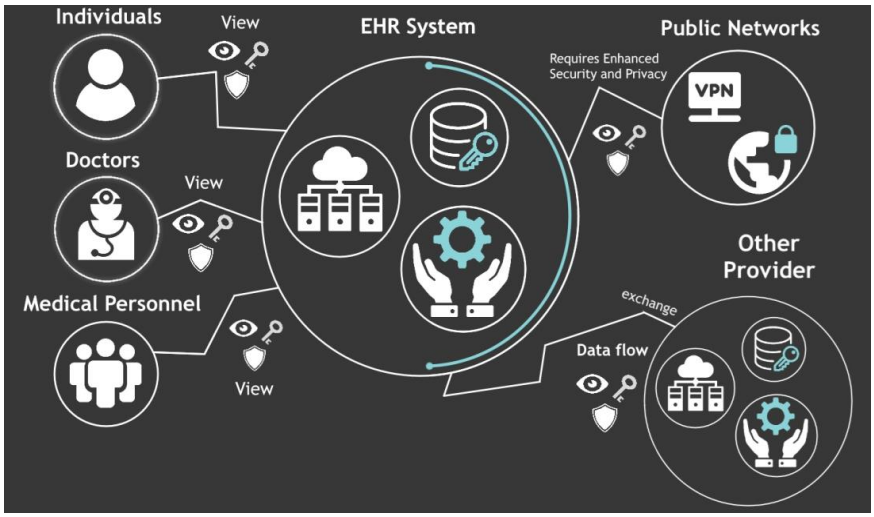

Figure 2. The EHR and the interchanges

Identification is possible in EHR, using other data sources and correlated datasets, which eventually help identify the entity. Profiling includes correlated information drawn from a number of sources which help predict the behaviour of the entity by means of the generated model. These profiles even if protected, can be used by marketing organizations serving advertising purposes or other causes, without providing the user with any notification or request for consent. It is a challenge to describe a model for giving consent directly bypassing any trusted entities. For describing how privacy properties apply in each user group differently, threat profiles have to be created. Extracting sufficient results in evaluation, requires specific challenges to be conducted. Every threat process conducted from different entity might be specifically analyzed. Other than that, specific methodologies for creating a threat model might apply. Actually, trusted entities might manage better and much secure these connections, rather in a P2P model, however this can be act different in a closed and private blockchain approach.

3 Towards A Decentralized Models: Blockchains

While there are great benefits from TTP-based models, there is also a growing public concern about transparency and user privacy. Updated services have evolved adding specific data disclosures offering facilitation for individuals to register easier to other services, using combined information. This way users share the same private information with many entities, without knowing the exact level of the required disclosure. Linked data offers great opportunities for data mining, but privacy is often

compromised. Moreover, third parties usually control or monitor network traffic and have access to personal information.

Third party trust is mandatory in order for third parties to be authorized concerning any information or data disclosures, keeping in mind that not only third parties are reliable. For example, most users trust personal information very easily in any third party. P2P approaches might not need third party trust while giving more control to the users. Blockchains might enhance the ability for linked data to exist, but can also be used in handling issues with regard to provision of consent, which is very important for protecting personal information, while at the same time ensuring integrity. Blockchains enable the realization of making transactions in a trust-less network without the need for validation by a central authority. Because of this feature, many decentralized applications have been developed, reaping the benefits that blockchain technology has to offer.

In contrary great concern is expressed about the impact of blockchains in privacy and about whether blockchains can be applied in privacy- preserving models. Major issues about performance, resource utilization, scalability and other issues are still in research. Summarizing, major issues of blockchains include significant computational power, energy and delay overhead. The impact of blockchains in privacy for the time being cannot be efficiently evaluated. Some of the differences of blockchains in contrast to TTP-based approaches are presented on Table 1.

TTP-based approaches	Blockchain as a Platform/Service
Privacy policies and security levels are become uniformed according to privacy and security policies	Scaling the levels of privacy, security and performance. Freedom for configuring privacy policies and security levels
Single point of control and failure	Scaled ownership and control over different distributed organizations, no single point of control or failure
Static policies and closed architecture	Control on inviting new members and configure different policies.
Usually in closed architecture, not always distributed, indirect	Open source, always distributed, direct
No special incentives	Incentives for running the distributed service and for "renting" sources costing efficiency

Table 1. Blockchains, decentralized applications and their impact

3.1 Private blockchains in Healthcare

In public blockchains transactions are shared through a shared ledger and are visible to all participants. Depending on the developed use case, transparency that blockchain offers might be restricted only to the enrolled participants. This can be achieved by using a mixing service, which enables a trusted third party to commit transactions instead, just like a "wallet proxy".

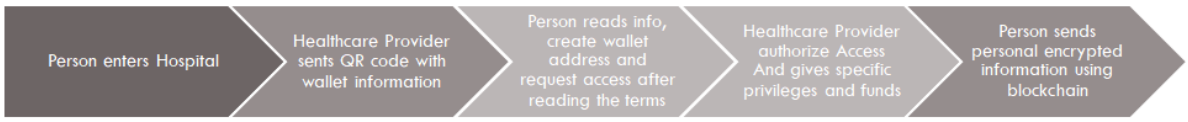

Figure 3. Use case for healthcare record management

The basic model in Figure 3, includes two different entities/nodes, the hospital and a Person who asks for access to the services. The Hospital has a QR Code for every person that enters. If consent is given, users are directed for fulfilling the required data. Digital consent is also handled by the blockchain. The system infrastructure using the blockchain also maintains the transactions. Specifically, the use case involves the handling of medical records using the blockchain technology. In healthcare providers, persons who enter the network include mostly patients. It is useful for the EHR System to have information and more specifically, the medical history of every member of the healthcare

institution. Having this information, institutes can provide the best service and also act faster in emergency situations.

For patients, having a full medical history and information about everyday habits, such as nutrition, sleep hours and other information, is very useful in modern approaches to healthcare management. In Healthcare, it is also crucial to have information about other personnel of the healthcare institute, including medical personnel and clinical visits. Modern Data Mining techniques developed techniques for extracting valuable data. Big Data Analytics can also enhance and improve the extracted results. However, handling personal information in the classic model can reveal many privacy issues. Since TTPs nowadays cannot be always trusted, especially in healthcare, more robust platforms have to be created with no single point of control and failure, giving more control to the users. Unapproved data disclosures are often happening in healthcare without the appropriate consent from the rightful holders. Decentralized approaches can offer this ability and in the same time "reward" the participants giving motivation for handling data to specific researches, while the process will be transparent. Keeping that in mind, privacy awareness and human behaviour is also critical for this approach to work appropriately.

3.2 Disclosure in Blockchains

We explained the infrastructure for the connection and data disclosure, but more work is needed for creating the application. The application will be a wallet manager, but also a useful data manager which holds personal information and data. The only user who will access this "account" or wallet, will be the one who knows the private wallet. The main concept is that the blockchain will hold the indexes of the actual data. For data disclosure, there are two options, even node C will enter the blockchain (Figure 4), with the required authorization, or node B will be entrusted for managing the disclosure (Figure 5,6).

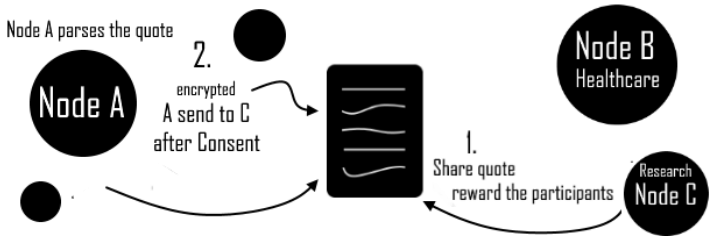

Figure 4. Direct consent and sending of information between node A and node C

In Figure 4, node C enters the blockchain and broadcasts a message. Node C as an entity will choose to share the needed quote in the blockchain. This quote will require anonymous information for specific criteria the research requires. Node C will ask for consent for every other node and will reward them accordingly. This option is a direct transaction from other nodes to node C.

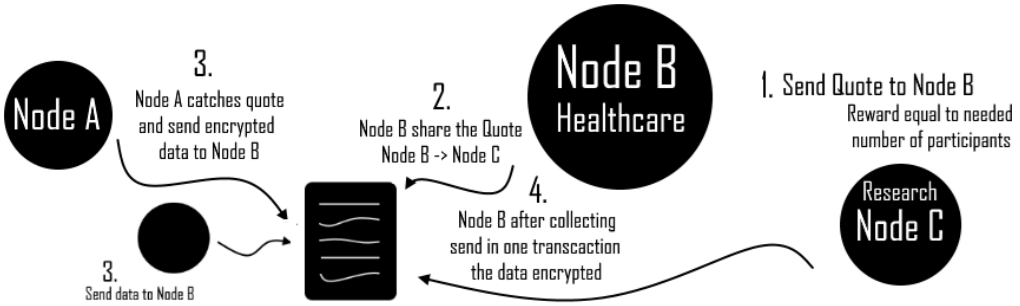

Figure 5. Node B is collecting data thorough the blockchain and sends it to node C

In Figure 5. Node B will be responsible for any required consent as also to handle the disclosure.

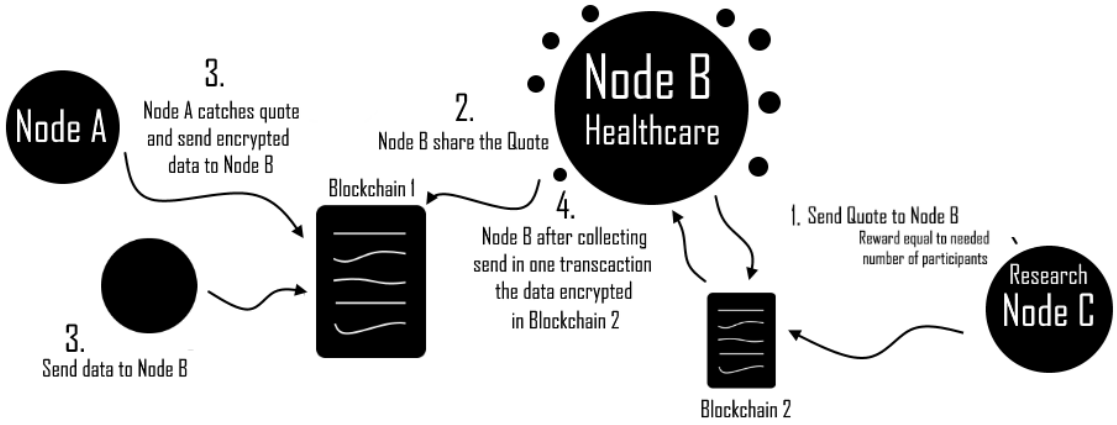

Figure 6. Node C exchanges data through a different blockchain.

Every node, participating in the specific disclosure will be rewarded. Node C will be able to see the transaction, even if it is encrypted. Node C, actually communicates with node B, knowing only the public wallet address of the healthcare provider. While the option in Figure 4, provides full control to the users, the model in Figure 5, enhances anonymity. In Figure 5, Node B is used as a mixing provider, which will send all of the data at once. A variation of this example, which includes holding transactions of Node B and Node C in a separate blockchain in Figure 6. Thus, Blockchain 1 is a private blockchain different from blockchain 2, giving more options for preserving the privacy properties.

3.3 Privacy and Security Analysis

In most cases, data is kept public in the blockchain. This may lead to concern over the issue of privacy. The data inside the blockchains might not be encrypted, making data completely public. Sending sensitive data through the blockchain requires encryption.

In the proposed use case, encryption is maintained, either by a local software or from the blockchain platform. In this specific use case, we manually execute the encryption, but the approach must have cryptography by default. However, enhancing privacy and security generates overhead and increases the amount of required resources.

Privacy Properties	Blockchains
Anonymity	On the proposed method in Figure 5 and 6, a mixing service will enhance this property. Using a mixing service will hide the identity of the sender along with other people.
Unlinkability	Unlinkability can be maintained by creating new addresses (shadow address) for a user. In this specific case, there are no measures for keeping Unlinkability intact unless this option is enabled. However, network analysis attacks, monitoring incoming and outgoing traffic can reveal that a single action came from the specific user.
Undetectability	Sending data along with funds will hide actual data through a large number of transactions. Randomly sending coins to new multiple addresses of the same owner make difficult for the adversary to recognize if these transactions are sent in random sequence or not.
Unobservability	If the transactions are private and mixed down, the adversary would not recognize dummy data from actual data. In private transactions, the adversary will only see “random” transactions from “random” mixing services.

Table 2. Privacy analysis

The main issue of blockchains is the possibility of attackers holding up to 51 % or more of computational power and resources inside the network. If this percentage occurs, the adversary will completely control the blockchain. Research is conducted for lowering this risk. Adding security measures to the blockchain will increase the overhead, with the possibility of adding extra latency to

the proof of work process and the tradeoffs of blockchains (Kiayias and Panagiotakos, 2015). More research about the performance of Bitcoin is conducted (Kogias et al., 2016), focused in the consensus latency.

Main issues appearing in blockchain technology, also apply in *Multichain*, concerning mining risks and lack of privacy. However, we proposed a more closed model by accepting specific users, authorizing them for accessing the blockchain. Zero knowledge proofs and blind signature protocol can be adopted for providing unlinkability. Privacy issues are maintained mostly by cryptography, but the linkability of transactions still exists. "Coin mixing" approaches can also be adopted and implemented by a trusted third party, but in a transparent way. Other approaches for preserving privacy properties are presented in Table 2 including the creation of "shadow wallets".

4 Conclusions

The total impact of blockchains in modern technologies has so far not been revealed. Performance issues including large amounts of system resources, vulnerabilities, as well as the privacy issues of blockchains have not yet been sufficiently evaluated. Privacy and security analysis has to be conducted in specific examples, in order for blockchain applications to be further developed. Developed platforms based on blockchains, including Ethereum, are very promising and more research has to be conducted for evaluating and evolving the application of blockchain technology in Information Systems. The full potential of the use of blockchains in the modern web, networks and information systems have not yet been revealed. Quantitative analysis while making complete experiments in specific network environments, can guarantee the adequate results about the impact of blockchains in privacy.

References

Agrawal, R., Kiernan, J., Srikant, R., & Xu, Y. (2002, August). *Hippocratic databases. In Proceedings of the 28th international conference on Very Large Data Bases* (pp. 143-154). VLDB Endowment.

Androulaki, E., Karame, G. O., Roeschlin, M., Scherer, T., & Capkun, S. (2013, April). *Evaluating user privacy in bitcoin. In International Conference on Financial Cryptography and Data Security* (pp. 34-51). Springer, Berlin, Heidelberg.

Buttyan, L., & Hubaux, J. P. (2007). *Security and cooperation in wireless networks: thwarting malicious and selfish behavior in the age of ubiquitous computing*. Cambridge University Press.

Ekblaw, A., Azaria, A., Halamka, J. D., & Lippman, A. (2016). *A Case Study for Blockchain in Healthcare:"MedRec" prototype for electronic health records and medical research data*.

Jacobovitz, O. (2016). *Blockchain for Identity Management.*

Kiayias, A., & Panagiotakos, G. (2015). *Speed-Security Tradeoffs in Blockchain Protocols. IACR Cryptology ePrint Archive*, 2015, 1019.

Kogias, E. K., Jovanovic, P., Gailly, N., Khoffi, I., Gasser, L., & Ford, B. (2016). *Enhancing bitcoin security and performance with strong consistency via collective signing*. In 25th USENIX Security Symposium (USENIX Security 16) (pp. 279-296). USENIX Association.

Kosba, A., Miller, A., Shi, E., Wen, Z., & Papamanthou, C. (2016, May). *Hawk: The blockchain model of cryptography and privacy-preserving smart contracts.* In Security and Privacy (SP), 2016 IEEE Symposium on (pp. 839-858). IEEE.

McConaghy, T., Marques, R., Müller, A., De Jonghe, D., McConaghy, T., McMullen, G., ... & Granzotto, A. (2016). *BigchainDB: A Scalable Blockchain Database*.

Miers, I., Garman, C., Green, M., & Rubin, A. D. (2013, May). *Zerocoin: Anonymous distributed e-cash from bitcoin. In Security and Privacy (SP)*, 2013 IEEE Symposium on (pp. 397-411). IEEE.

Modi, C. N., & Patil, A. R. (2016). *Privacy Preserving Association Rule Mining in Horizontally Partitioned Databases Without Involving Trusted Third Party (TTP). In Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics (pp. 549-555). Springer India.*

Pfitzmann, A. and Köhntopp, M. (2017). *Anonymity, Unobservability, and Pseudonymity — A Proposal for Terminology.*

Pfitzmann, A., & Hansen, M. (2010). *A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management.*

Rachovitsa, A. (2016). *Engineering and lawyering privacy by design: understanding online privacy both as a technical and an international human rights issue.* International Journal of Law and Information Technology, 24(4), 374-399.

Ranchal, R., Bhargava, B., Othmane, L. B., Lilien, L., Kim, A., Kang, M., & Linderman, M. (2010, October). *Protection of identity information in cloud computing without trusted third party.* In Reliable Distributed Systems, 2010 29th IEEE Symposium on (pp. 368-372). IEEE.

Schanzenbach, M., & Banse, C. (2016, September). *Managing and Presenting User Attributes over a Decentralized Secure Name System.* In International Workshop on Data Privacy Management (pp. 213-220). Springer International Publishing.

Zyskind, G., Nathan, O., & Pentland, A. (2015). *Enigma: Decentralized computation platform with guaranteed privacy.* arXiv preprint arXiv:1506.03471.