Using Blockchain to Secure Digital Identity and Privacy Across Digital Sectors

Jaynill Gopal and Stacey Omeleze-Baror
Department of Computer Science, EBIT, University of Pretoria, South Africa

jgopal@tuks.co.za
Stacey.baror@cs.up.ac.za

Abstract: In the digital age digital data has been seen as the new currency for both companies and bad actors, leading to mismanagement of personal data by both entities. This mismanagement could lead to personal data being breached while this may prove to be beneficial to certain entities, this however is not the case for users and digital citizens. In recent years there has been an influx in data breaches and data mismanagement cases throughout various industries, including corporations such as Apple and Facebook, involving critical data relating to user's digital identities, personal data, and other identifiable information. This issue may be addressed by employing the use of blockchain technology, a technology that has been recognized as a secure system promoting security and privacy, we can prevent this mismanagement of data and data breaches from happening. While blockchain is a relatively new technology, there is potential to use it in the current age of the internet and digital identities by employing blockchain technology to secure one's digital identity. This research aims to propose a potential solution to the issue of data protection and data breaches by proposing and making use of a conceptual model which makes use of a multi-level blockchain system. The system isolates data from digital identities between various digital platforms to minimize the amount of data breaches that may occur – further minimizing the amount of personal data which may be breached. This system allows the user to have full control over who may access their private data, while preventing bad actors from accessing the data without the user's authorization. The multi-level blockchain splits the user's data according to industrial or digital sector in which their data is used, with a master blockchain acting as the connecting link to a person's digital identity. Making use of this multi-level blockchain allows for the user to control who has access to their data, while remaining anonymous and secure in a digital platform's database or digital storage.

Keywords: Digital identity, Blockchain, Data protection, Digital privacy, Privacy, Data breach

1. Introduction

The introduction of the internet to the public has led to many advantages with many people becoming accustomed to the versatility of the internet, with popular uses to the average user being social media, online banking, and shopping. While the internet makes the life of the average user easier, most share their information willingly online (contact details, and other identifiable information) and some of this information is repeated across those various platforms. The user believes that this information is kept safe and confidential by these online platforms, however there are threats to users' online identity which they are often unaware of, the greatest threat being data breaches (Neto et. al., 2021). A data breach is an event where an unauthorised entity gains access to user or client data stored in a digital platform's database, these events may be due to either internal or external entities (Mohammed & Tejay, 2023). A data breach could also be defined as data that is sold by those platforms. This threat could be dangerous to user data in the wrong hands, these wrong hands are not limited to just cyber attackers, but they also refer to those who use that information. Once this data is in the hands of malicious parties (or third parties), a user's digital identity is exposed and can be used to identify a user and target that user directly – either through marketing, phishing attacks, account high jacking, etc. While most of these may seem harmless, there could be a harmful impact to the user should their digital identity be made public to those unauthorised parties.

Since most users use the same login credentials to login to different digital platforms (Gao, et. al. 2020), this information can be considered as part of their digital identity and identifiable information. By protecting this information, we would be able to prevent targeted attacks on a user's digital identity. This paper introduces a method for protecting a user's digital identity using blockchain, a technology which is used to protect data and anonymise its users. Aiming to protect a user's digital identity by keeping the user anonymous between different digital platforms, while still being able to be identified by the platform as an anonymous digital identity.

2. Background Information

Blockchain technology, since its release in 2009, has been recognised for its security and reliability through its decentralised nature. Blockchain brings forth confidentiality, and privacy by using cryptography to secure the data which it contains. By replicating its data to nodes in a blockchain network, blockchain can achieve a

decentralised nature by bringing a reliability to the system – being resistant to disasters which centralised systems suffer from. In the case of blockchain if a node in the system goes offline, there is another available to serve the network (Dai, et. al., 2017). The architecture of blockchain consists of a chain of blocks that are stored in a node, a block stores data in the system, and a node is a computer system which holds a copy of the blockchain (Zhang, Xue, & Liu, 2019). Security in blockchain comes from the cryptography and immutability properties of the system, the decentralised nature of blockchain requires that all nodes need to have the same information. This is achieved by a consensus mechanism that blockchain employs, if one node differs from others then the node will be deemed invalid (Mohsin, et al., 2019). While blockchain is an attractive technology in terms of security, there are however drawbacks namely: storage, computability, and energy consumption. These drawbacks, depending on the use case, can be negligible if security is the main aim of the system that the blockchain is being used in. With privacy and security being one of the focus points of blockchain technology, anonymity is enforced – highlighting the privacy aspects of the blockchain, this is because blockchain uses hashed identifiers to store identities (Takemiya & Vanieiev, 2018).

A digital identity is an identifier which is used to identify a user over a computer network (Ante, Fischer, & Strehle, 2022). A user's digital identity contains information such as a user's login name, personal details, location information and other information depending on the platform for which the identity is hosted on (Saklikar & Saha, 2006). In recent years sharing personal information on digital platforms have become ever prevalent, this prevalence has brought a concern that users often do not consider the reprecussions of sharing their information online – they simply click on accept when digital platforms ask the user to review and accept the terms of using that specific platform without reviewing those terms (Chan & Virkki, 2014). This behaviour has been seen since the rise of social media and as the internet gained popularity in the past decade, this negligence for digital privacy has made it easier for digital platforms to process and analyse data without the user's knowledge. Processing of user data does not need to be done soley by the digital platform, as is common practice by digital platforms the user data collected is often sold to third parties who may or may not have malicous intentions (Saglam, Nurse, & Hodges, 2022).

The recent rise in reported data breaches over the past decade, with over 100 breaches being reported per year, has brought on a concern for information and data protection by netizens and government regulators alike. A famous breach being the 2014 iCloud breach, where private pictures of celebrities were leaked to the public because of this breach (BBC News, 2014). While this breach was not through an attack on Apple's iCloud infrastructure, but rather reported as targeted attack on the users' accounts by gaining unauthorised access using the victims' login credentials to gain access to those pictures. This unauthorised access was done due to the attacker acquiring the login ID of their intended victim, then by either brute forcing the victims' passwords, or using their security question answers to gain access to the accounts (Marwick, 2017). While preventing brute forcing passwords and answering security questions may be preventable or unavoidable, we can protect a user's login ID by protecting their digital identity.

Another example of a data breach is the Facebook Cambridge Analytica data scandal which came to light in 2018. The focus of this scandal was due to Facebook building up psychological profiles based on its users' interactions on and off the platform. These psychological profiles were then used against the users in the form of targeted advertisements (Hinds, Williams, & Joinson, 2020), while this is common practice in the digital realm, the data that Facebook obtained from those users', psychological profiles and other identifiable data were handed off to a third party – Cambridge Analytica (CA) – without the users' consent (Shipman & Marshall, 2020). In this case, CA was not bound by Facebook's terms and conditions agreement with Facebook's users and were seemingly free to do what they wanted with the data that Facebook had handed to CA. This form of data usage is common practice in big data (Hashimova, 2016) and is used in targeted digital advertising, this is due to advertisers being able to link a user's digital identity to an advertisement that is based on a user's digital history.

While there are authentication methods which do allow login using third-party platform's single-sign on (SSO) and OAuth which limit the need for resharing personal information (Sadqi, Belfaik, & Safi, 2020), there are some security and privacy concerns using SSO (Karie, et. al., 2020). The information shared to the platform using SSO is not in the user's control, using additional trackers attached to the user's session (Pham et. al., 2023).

3. Methodology

This paper introduces the Multi-layered blockchain (MLB) system model to act as a guide to protect users' digital identities on the internet. The methodology used in the proposed model would be a combination of

literature reviews pertaining to the research topic, the viability of blockchain usage in different industries in terms of security and data privacy to one's digital identity. The result being a model of a multi-tier blockchain system to facilitate the storage and separation of data to protect one's digital identity.

3.1 Multi-Layered Blockchain System Model

This model would demonstrate that a user would be the sole possessor of their information, with a sub-blockchain created for each digital platform that a person chooses to share their digital identity with. The blockchain effectively forms a pseudo hash-table using blockchain with each node being related to a digital platform that a user chooses to share their information with, and only have access to that node.

Each digital platform that a user has a profile on will have an internal blockchain of user profiles which contain a link to an identity in the Global blockchain (GB), the digital platform's blockchain would not contain personal information associated with the user's digital identity - only the user's account information associated with the digital platform would be stored. This would encapsulate and isolate important digital platform and user information, preventing access to that information outside of the digital platform. The global blockchain would be the basis of authenticating a user against a specific digital platform profile.

Figure 1 provides a visual representation of the MLB. As seen in GB, there are digital platform chains which connect to an identity. Each identity is stored in a separate blockchain and is logically separated from the digital platform that the user has an account with. The User Identity Blockchain stores personal and identifiable information which is often asked by websites, such as a user's name, contact information, and other identifiable information.

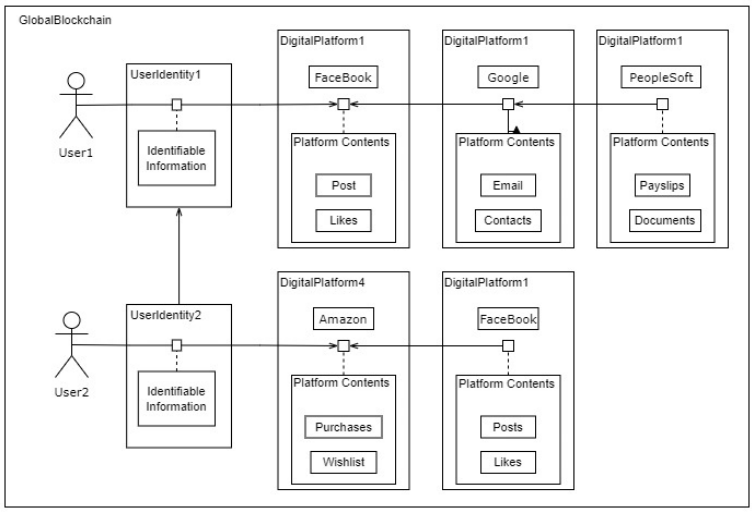

Figure 1: Visual Representation of the Multi-level Blockchain System

Figure 2 represents the internal blockchain that will be used by a digital platform. As seen each user profile has a link to a user identity within the global blockchain as depicted by the blocks *UserIdentity*. As seen in the *UserProfile* blocks, the information stored in the user profiles are only that of which is relevant to the digital platform. All non-platform specific user information is stored in the user identity.

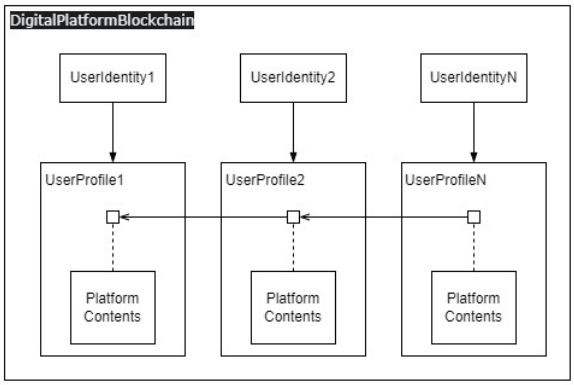

Figure 2: Visual Representation of the Digital Platform blockchain

3.2 Specification

The design of the MLB is a, as the name suggests, blockchain system with multiple blockchains. As mentioned in the previous section, the system consists of a Global blockchain, and an Internal blockchain. This approach provides two advantages – to prevent data duplication across digital platforms, and to ensure that should a data breach or leak on a digital platform occur the personal information of its users are not compromised in the breach. This is done by isolating the digital platform from its users, allowing the user a single place to store their digital information as well as to prevent attacks on digital platforms by attackers trying to exploit vulnerabilities on a digital platform to gain access to its user information.

3.2.1 Global blockchain

The Global blockchain is a public blockchain which acts as the registrar of all Digital Identities, where a single block in the system consisting of a single user's digital identity; since each user will have their own block, each block will contain a unique identity. The blockchain will use a hybrid technique depending on the operational scenario that would require the Global Blockchain. New blocks will require a Proof of Stake in order to be added to the chain, this will only occur when a user first creates their digital identity. New nodes will require Proof of Work, Proof of Authority, and Delegated Proof of Stake – this is to ensure that the node is of a trusted source. Transactions relating to specific blocks such as requests for information will require a Proof of Authority in order to validate the request and authenticate it accordingly.

A digital identity block would contain personal information about a user that is shared with a digital platform, information such as a user's contact information, geographic information, demographic information, financial information, or any other information which a user is likely to share to a digital platform. When a digital platform requests the user to access the information in the form of a claim, a request is sent to the user for the information. The user will need to review and authenticate the request before the information is sent to the digital platform in the form of a payload, which contains only the information that is requested by the digital platform. The full contents of an identity block will not be released to the digital platform, only the contents recognised by the claim that the digital platform has requested will be sent.

Requests for information are to be requested from a digital platform that has a valid claim granted by a trusted authority, unsigned or self-signed claims will not be accepted by the Global blockchain to be sent to the user for approval. Requests from an unregistered, or untrusted platform will be automatically denied by the Global Platform. Should a claim be successful and approved by the user, the requested information will be sent in an immutable format, with an appropriate time to live limitation. This is to prevent the digital platform from storing the information indefinitely as well as to prevent the information from being altered or distributed.

3.2.2 Internal blockchain

The Internal Blockchain is a private blockchain which is specific to the digital platform, this means that the blockchain may be implemented however the platform sees fit. The contents of the blockchain would consist of blocks containing information that is specific to the digital platform. Each block would contain a reference to the associated digital identity block in the Global blockchain, see Figure 2, as well as platform specific information. Since there is a reference to a user's digital identity in the Global blockchain, the internal blockchain does not need to – and should not – contain any personal information related to the user. If the platform requires the use of a user's personal information – such as the user's cell phone number – the platform should request that information via a claim.

A digital platform should register to apply for specific claims (depending on the category which the digital platform falls a part of) from a trusted authority, a digital platform can only apply for a claim which requests platform specific information – a social media platform may not request a user's credit card information. Should a digital platform request information via the claim and the request is approved, the platform will have a limited period to process that information before the authorisation would be revoked and the information will be discarded from the platform.

4. A use Case for the Multi-Layered Blockchain System

Suppose we have User1 (U1) who creates a profile on a digital platform (DP1) using the conventional methods of profile creation on digital platforms. U1 shares their name, email address, cell phone number and their social security number on the platform. The user uses the platform for a few months, soon after they discover that they are receiving many spam emails and phone calls than usual. The user is annoyed by this occurrence;

however, they think nothing of it. A few days later, they learn that DP1 has suffered a data breach where all account information of all its users including contact details and social security numbers, have been leaked a few days after U1 had created their account. After finding out about this incident, U1 decides to check if their social security number had been used after the leak and found that it is now associated with several credit accounts which U1 has no knowledge of. This mean that U1 would need to undo the damage caused by the entity or entities using U1's details, which could be a lengthy process. In this scenario the data breach had led to U1's identity being stolen, and their contact information being sold, as a result of the data breach on DP1.

The Multi-layered Blockchain system (MLB) aims to prevent this from occurring, suppose that the system is in use on a global scale. The global blockchain (GB) would contain all personal data needed for a user's digital identity, including their social security number, and contact details. Suppose we have User2 (U2) who created a profile on digital platform2 (DP2) using the MLB. The user would create their profile on DP2 by linking their digital identity to DP2. DP2 will only have access to the user's digital identity, and not the contents of the identity, Figure 3 illustrates how the processes used in the MLB.

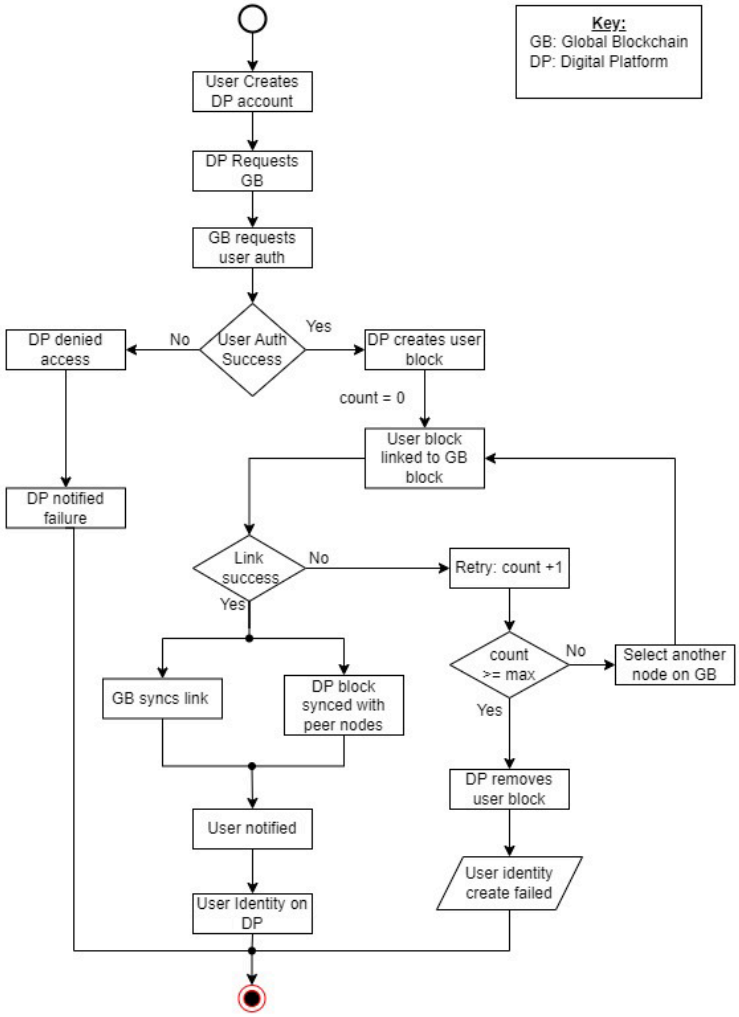

Figure 3: Process used in the Multi-layered blockchain

Once U2 chooses to link their digital identity to DP2, DP2 would send a request to the GB, which would request the user to approve this request to link their identity to DP2. Once this request has been approved by the user, U2's digital identity is subsequently linked to their profile on DP2. This link to DP2 will merely form as a link to the user, and DP2 will not have access to any of U2's personal information, Figure 4 represents the logical flow of the process involved should a digital platform request to access a user's information.

After some time, DP2 suffers from a similar data breach that DP1 had suffered from. In this incident however the data that had been breached was a list of profiles on DP2, with no accessible personal information to U2 – only the hash code containing the address of U2's digital identity on the GB was exposed. Sometime after the breach U2 has noticed a request to access information by another actor, which U2 has no recollection of

linking their digital identity to, U2 decides to deny this request for information as well as block any requests coming from this actor. This request will be denied by the Global blockchain, and any subsequent requests will result in the GB recognising them as an invalid request. In this incident, U2's information was protected by the MLB and as a result they did not suffer the same fate that U1 had suffered from.

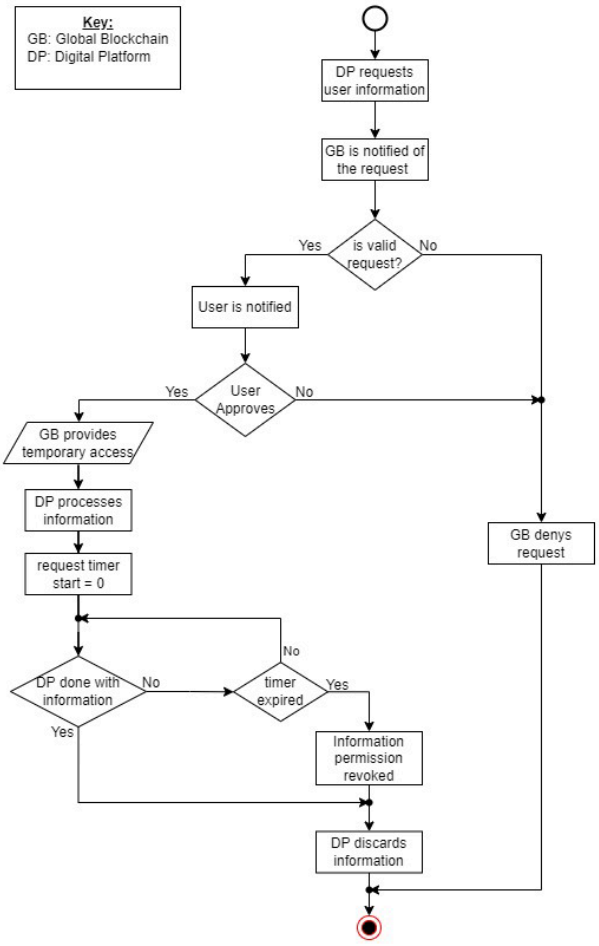

Figure 4: Request for information process using the Multi-layered Blockchain

5. Evaluation

The MLB model provides the user with a protected digital identity through storing their digital identity in a blockchain, while preventing data duplication between different digital platforms and risking their data from being compromised through data breaches on digital platforms. This allows the user's identity to be isolated from a digital platform, and any information stored about the user on is safely stored on the platform's internal blockchain. The evaluation on this approach will be conducted as follows: advantages, disadvantages, and considerations.

5.1 Advantages

The advantages of the MLB are that the use of a blockchain to store a user's digital identity inherits the advantages of a blockchain. The key advantages being that the system protects the data from being altered, prevents unauthorised access, the data is stored in a decentralised manner, and most importantly transparency and security (Naik & Jenkins, 2020). The use of multiple blockchains between the user and the digital platform, allows both the isolation of the user's digital identity, and pseudo-anonymisation on the platform. This isolation prevents the user's information being compromised on the digital platform due to the information not being stored on the platform, this will also reduce the number of user information motivated attacks on the platform. Should there be an attack on the digital platform, the information stored would be stored on the internal blockchain – making the data gathered useless unless the attackers have the time and resources to decrypt the contents of the blockchain.

The use of claims to allow digital platforms to request access to information from the user, along with a time-to-live limitation on the data controls the information that the digital platform is allowed to access. This restriction prevents the digital platform from storing the user's information indefinitely and unsecured, as well as prevents the platform from distributing the data to third parties without the user's consent. The use of claim requests also makes the user aware of where their information is being used and the purpose for requesting their information, the user is then given full authority to deny or allow the request should they feel fit.

5.2 Disadvantages

As mentioned in the previous section, the MLB inherits the traditional properties of blockchain technology, this means that the MLB inherits blockchain's disadvantages as well. Due to the decentralised nature of blockchain, if either the Global Blockchain or Internal Blockchain is compromised, then the data stored in the respective blockchain will be compromised (Dai, et. al., 2017), while this will not be an issue on the digital platform's internal blockchain regarding user data, this is of concern on the Global Blockchain due to the user identities being stored on the blockchain, future work on this research will attempt to prevent this from occurring.

The energy consumption and scalability of the MLB is another aspect to consider, while this is not much of a concern on the Global Blockchain, it is a concern on the digital platform's internal blockchain, depending on the frequency of use that the digital platform has. The Global Blockchain is not of concern regarding scalability and energy consumption when generating nodes, as the number of nodes being directly proportional to the population of the digital space, i.e. one node per digital identity and only one digital identity can be associated with a person in the world. This means that on initial creation of the Global Blockchain, energy consumption will be high. However, once this process is complete consumption will be significantly reduced when a digital identity is added after this process.

5.3 Considerations

While there are advantages and disadvantages of the MLB, there are considerations that need to be accounted for regarding the authority of the system. The authority that controls the claims in the system will need to be an entity that is trusted by all parties involved, being impartial, with the integrity to assign and sign claims for digital platforms. This is to ensure that user trust is enforced between the digital platform and that the user feels safe that their information is not being used for purposes other than what it was intended for.

The second consideration regarding authority is that the global blockchain needs to be a private blockchain. This is to prevent the chain from being compromised by a syndicate of malicious nodes as well as prevent the blocks' data on the blockchain from being readily available to the public or malicious actors. A block's data can include the historical data of the chain, headers, data, and importantly peer information. By limiting the nodes to trusted entities, we can prevent peer information from being available to malicious actors, preventing attacks on other nodes in the blockchain.

An important consideration is the impact that the MLB would have on current authentication schemes as well as the architectural changes that would be needed to implement the MLB on a global scale.

6. Conclusion and Future work

This paper introduces a model of the Multi-level Blockchain (MLB) a system which consists of two blockchains which aims to ensure user privacy and prevent user information from being compromised when data breaches occur, while also preventing user data duplication between digital platforms from occurring. Consisting of a Global Blockchain containing users' digital identities, and digital platforms Internal Blockchain containing platform specific information, the MLB can achieve this aim by both securing a user's digital identity and isolating their identity from a digital platform.

The next steps pertaining to this model is to implement the model programmatically and experimenting further by simulating a working example of this model. This would be able to examine the system and bring forth any limitations or pitfalls which this system may occur that has not been discussed in this paper as well as a comparison of the MLB compared to existing authentication schemes. Future work will also research the aspects of the system mentioned in the evaluation section, bringing into consideration the disadvantages, and determining methods to minimise these disadvantages.

References

Ante, L., Fischer, C., & Strehle, E. (2022). A bibliometric review of research on digital identity: Research streams, influential works and future research paths. *Journal of Manufacturing Systems, 62*, 523 - 538. doi:https://doi.org/10.1016/j.jmsy.2022.01.005

BBC News. (2014, 09 2). *Apple confirms accounts compromised but denies security breach*. Retrieved September 19, 2023, from BBC News: https://www.bbc.com/news/technology-29039294

Chan, C., & Virkki, J. (2014, January). Perspectives for Sharing Personal Information on Online Social Networks. *Social Networking, 3*. doi:10.4236/sn.2014.31005

Dai, F., Shi, Y., Meng, N., Wei, L., & Ye, Z. (2017). From Bitcoin to cybersecurity: A comparative study of blockchain application and security issues. *2017 4th International Conference on Systems and Informatics (ICSAI)*, (pp. 975-979). doi:10.1109/ICSAI.2017.8248427

Gao, X., Yu, L., He, H., Wang, X., & Wang, Y. (2020). A research of security in website account binding. *Journal of Information Security and Applications, 51*. doi:10.1016/j.jisa.2019.102444

Hashimova, K. (2016, 07). The Role of Big Data in Internet Advertising Problem Solution. *International Journal of Education and Management Engineering, 6*, 10 - 19. doi:10.5815/ijeme.2016.04.02

Hinds, J., Williams, E. J., & Joinson, A. N. (2020). "It wouldn't happen to me": Privacy concerns and perspectives following the Cambridge Analytica scandal. *International Journal of Human-Computer Studies, 143*.

Karie, N. M., Kebande, V. R., Ikuesan, R. A., Sookhak, M., & Venter, H. S. (2020). Hardening SAML by Integrating SSO and Multi-Factor Authentication (MFA) in the Cloud. *Proceedings of the 3rd International Conference on Networking, Information Systems & Security, NISS '20*(56), 6. doi:10.1145/3386723.3387875

Marwick, A. E. (2017, 09). Scandal or sex crime? Gendered privacy and the celebrity nude photo leaks. *Ethics and Information Technology, 19*(3), 191. doi:10.1007/s10676-017-9431-7

Mohammed, Z. A., & Tejay, G. P. (2023, February). How to Compensate After a Data Breach? Investigating Compensation Types and Role of Fairness in Customer Repatronage Intentions. *SIGMIS Database, 54*, 110–127. doi:10.1145/3583581.3583588

Mohsin, A. H., Zaidan, A. A., Zaidan, B. B., Albahri, O. S., Albahri, A. S., Alsalem, M. A., & Mohammed, K. I. (2019, May). Blockchain Authentication of Network Applications: Taxonomy, Classification, Capabilities, Open Challenges, Motivations, Recommendations and Future Directions. *Comput. Stand. Interfaces, 64*, 41–60. doi:10.1016/j.csi.2018.12.002

Naik, N., & Jenkins, P. (2020). Governing Principles of Self-Sovereign Identity Applied to Blockchain Enabled Privacy Preserving Identity Management Systems. *2020 IEEE International Symposium on Systems Engineering (ISSE)*, (pp. 1-6). doi:10.1109/ISSE49799.2020.9272212

Neto, N. N., Madnick, S., Paula, A. M., & Borges, N. M. (2021, January). Developing a Global Data Breach Database and the Challenges Encountered. *J. Data and Information Quality, 13*. doi:10.1145/3439873

Pham, T.-H., Vo, Q.-H., Dao, H., & Fukuda, K. (2023). SSOLogin: A framework for automated web privacy measurement with SSO logins. *Proceedings of the 18th Asian Internet Engineering Conference, AINTEC '23*, 69–77. doi:10.1145/3630590.3630599

Sadqi, Y., Belfaik, Y., & Safi, S. (2020). Web OAuth-Based SSO Systems Security. *Proceedings of the 3rd International Conference on Networking, Information Systems \& Security, NISS '20*(69), 7. doi:10.1145/3386723.3387888

Saglam, R. B., Nurse, J. R., & Hodges, D. (2022). Personal information: Perceptions, types and evolution. *Journal of Information Security and Applications, 66*, 103163. doi:https://doi.org/10.1016/j.jisa.2022.103163

Saklikar, S., & Saha, S. (2006). User Privacy-Preserving Identity Data Dependencies. *Proceedings of the Second ACM Workshop on Digital Identity Management* (pp. 45 - 54). Alexandria, USA: Association for Computing Machinery. doi:10.1145/1179529.1179537

Shipman, F. M., & Marshall, C. C. (2020). Ownership, Privacy, and Control in the Wake of Cambridge Analytica: The Relationship between Attitudes and Awareness. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1 - 12). New York, NY, USA: Association for Computing Machinery. doi:10.1145/3313831.3376662

Takemiya, M., & Vanieiev, B. (2018). Sora Identity: Secure, Digital Identity on the Blockchain. *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, *02*, pp. 582-587. doi:10.1109/COMPSAC.2018.10299

Zhang, R., Xue, R., & Liu, L. (2019, July). Security and Privacy on Blockchain. *ACM Comput. Surv., 52*. doi:10.1145/3316481