Imperial College London

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

Decentralised Digital Identity Management: A Mobile-Friendly System on Blockchain

Author:
Edward Jenks

Supervisor:
Dr Thomas Lancaster

Submitted in partial fulfillment of the requirements for the MSc degree in MSc Computing of Imperial College London

September 2022

Abstract

As the UK moves towards establishing a framework for a secure and trusted digital identity [1], an opportunity for establishing a new, decentralised paradigm with unparalleled privacy and security for users presents itself. This report details the design, development, and implementation of a General Data Protection Regulation (GDPR) compliant [2], blockchain-based, biometric identity management system and a digital proof-of-age application that interacts with it. In doing so, a siamese image recognition neural network [3] capable of encoding biometric face data within a 64-byte output was developed through appropriate architecture selection and experimentation. This output is small enough to be compatible with storage on blockchain [4]. An anonymous certification system that allowed user data to be verified and was capable of preventing duplicate account creation was also designed to ensure that the blockchain-based data storage would be complaint with GDPR. Finally, the system and accompanying app were developed with care taken to ensure identity verifications remained free, fast, and reliable by trials of multiple blockchain topologies and smart contract designs. The result of the project is a highly feasible and effective system that can be integrated into a mobile application in such a way that it requires no technical knowledge or additional steps to use. The design is portable and can be extended to many use cases, which is essential for enabling consistent and secure digital identity in the future.

Acknowledgments

This project was completed under the supervision and guidance of Dr Thomas Lancaster.

Contents

Chapter 1

Introduction

Since the introduction of blockchain technology in 2008 [5], there has been vast developments in the space as new applications have been identified. What began as an unregulated, decentralised currency has now developed into a sector where issues of intellectual property ownership [6], digital exchange of goods [7], and video games [8] have seen decentralised implementations. One possible application of blockchain technology is digitised identity management, a system for identifying individuals. This has been attempted multiple times and remains a challenging issue to square with the core values of decentralisation.

At its inception, cryptocurrency was designed to avoid control exerted by governments and large non-government entities. This ideal is hard to bring into the real world where regulations have many benefits. The reluctance of cryptocurrencies and decentralised projects to work alongside governments and other regulating bodies is partly responsible for the lack of widespread adoption of blockchain technology. It instead remains a volatile investment platform, topic for enthusiasts, and a hotbed for money laundering and criminal activity. For users to experience the benefits of blockchain and decentralisation, namely the privacy, transparency, and security it can provide, steps need to be taken to evaluate how the technology can be integrated into our existing systems and laws.

1.1 Project Aim

In this report, the development and evaluation of a decentralised biometric identity management solution that conforms to current UK legislation is presented. The integration of biometrics into this identity solution provides it with a ground truth to uniquely distinguish individuals without sole reliance on central authority or preassigned identity documentation. A demonstration of how this system can be used in the real world is given with a digital proof-of-age mobile application. This would allow users to buy age-controlled goods with their mobile phones in a format compliant with the upcoming support for digital IDs in the UK. The digital proof-of-age app demonstrates rapid peer-to-peer (P2P) verifications where a user can reliably and securely identify themselves to another party without exposing personal data to anyone else.

An identity management solution requires two main functions. Firstly, users must be able to apply for a place within the system by proving their identity. Secondly, users must be able to prove their identity to others using the system. For the purpose of this report, the former will be referred to as authentication, and the latter will be referred to as verification. Both of these functions are incorporated in the proposed design.

An important part of the authentication process is confirming that a potential user matches the identification document they have provided. This process must conform to the standards laid out in the Government's Good Practice Guide (GPG) 45 [9]. Typical systems involve submission of a picture or video of the user's face along with photos of their document. A combination of human and neural network inspection is then used to check for the match [10]. Since this part of the authentication process is strictly governed and has well-established solutions, it will not be covered within the scope of this project. It is likely that a real world implementation would take advantage of one of the many government-approved third party services to carry out this task [11]. Instead, this project focuses on other aspects of the authentication process that tackle forms of fraud made possible through the anonymity of accounts certified on the blockchain, and on the security of a decentralised solution.

1.2 Identity Management Systems

In the development of an identity management system, it is important to understand the risks and challenges involved in the area. These are related to the experience of the users within the system and the opportunity for bad actors to commit acts of identity fraud. To ensure these issues are understood, they have been briefly summarised below.

The Laws of Identity

In 2005, Microsoft's Kim Cameron laid out her 'Laws of Identity' [12]. These were a set of criteria for digital identity to meet in the context of trying to establish an internet identity layer. The laws have formed a strongly defined foundation on which the issue of digital identity has been further discussed in the years since. They perfectly capture the importance of users' experiences and rights within a digital identity system. They are as follows:

1. **User control and consent:** The system must be designed in a way that keeps users in control.
2. **Minimal disclosure for a constrained use:** The solution that uses the least possible information and discloses as little identifying data as possible is the most sustainable.
3. **Justifiable parties:** The system must make the user aware of any data processing and inform them of the parties the data is shared with.

4. **Directed identity:** An identity system must support both omnidirectional and unidirectional identifiers.

5. **Pluralism of operators and technologies:** A perfect system would rely on multiple providers and not be bound to a single technological form.

6. **Human integration:** Human behaviour must be factored into the design from the ground up.

7. **Consistent experience across contexts:** A digital identity can be used in many different contexts. One should seek to consolidate these contexts into a single, continuous experience.

First and foremost, she highlights the importance of prioritising user control and consent pertaining to their data. This is core to the development of this project, and central in every decision made. The concept of 'least identifying information' is used to always minimise a user's exposure by using the smallest amount of data possible to achieve a given task. Thought should be given as to how this relates to the verification process of this project's digital identity system. Another key 'law' is human integration, illuminating the importance of designing systems in a human-centered manner. For a digital identity management system to be adopted, it requires general acceptance. It is therefore vital to make the system as easy to use as possible, removing all technical aspects of dealing with blockchain transactions from the users' concern.

Identity Fraud

There are four main forms of identity fraud related to a proof-of-identity. A digital identity management system must consider these in two ways. Firstly, it must be designed to combat users trying to go through the authentication process with a fraudulent identity of some kind. Secondly, it must combat users attempting to commit fraud within the identity management system. The forms of proof-of-identity fraud are given below:

- **Stolen Genuine Identity**: This form of fraud occurs when a document or credential that proves an individual's identity is stolen and used by another party without their consent [13].

- **Shared Genuine Identity**: This form of fraud occurs when a document or credential is used by an individual that it does not belong to with the permission of the owner of the identity [14].

- **Modified Genuine Identity**: This form of fraud occurs when the fraudster gains access to a genuine credential or document and modifies it for their use [15].

- **Fake Identity**: This form of fraud occurs when a fraudster makes a false document or credential [15].

Identity fraud is a serious issue in the UK. Nearly 500 identities are stolen daily [16] and it makes up 53% of online fraud [17]. Fake identities can cost as little as £10 [18]. Digitisation opens up new avenues for fraud with the relative ease of copying, modifying, or intercepting data, even when measures to protect it have been put in place. However, it also offers an opportunity to innovate and develop new, more secure identity solutions.

1.3 Background Work

Digital Identity Management

Past research on digital identity management has largely focused on the development of an internet identity layer [19, 20]. This is a space where decentralised technology can be applied with relative ease to provide a solution that protects user data and anonymity. However, it comes without the burden of legislation surrounding government-approved identifications. The identity management solution developed in this project has applications such as proof-of-age, credit scoring/allocations, banking and electoral registration in mind. It must therefore meet a much stricter set of criteria to fulfill its desired purpose. In doing so, some degree of the decentralisation will be lost. This is the cost of the government itself being a centralised entity that requires some degree of accountability in the case of failure or vulnerability in the system. Additionally, they require assurance that the integrity of the system meets their minimum requirements. In this project, instead of ignoring these facts in hope of a Utopian future where decentralised technology is accepted unilaterally, these aspects are embraced and built in from the beginning to give users the best possible practical outcome.

Blockchain and Legislation

The primary legislative concern of this identity management system is the General Data Protection Regulation (GDPR) [2], the UK's data protection and privacy rights. In particular, GDPR enshrines a 'right to erasure', which requires a service to delete all personal data belonging to a user on request, whether it is encrypted or not [21]. As blockchain is immutable, it is not possible to delete data that has been stored on it, as a footprint will always be present in the past blocks of the network [5]. No system up to this point has been developed with this in mind. It places an important limitation on the design of the identity management solution. Personal data cannot be stored on the blockchain. Instead, some certification that can verify personal data reported by a user to be correct must be used. A solution of this form introduces many secondary benefits along with some of its own challenges. Namely, it makes all users on the blockchain completely anonymous. While this is good for privacy, it makes governance of the system and prevention of fraud challenging. Therefore, a complete biometric authentication process that tackles these issues will be outlined and tested.

Biometric Identification

Biometric identification is the use of biometric markers to distinguish between individuals [22]. A wide set of biometric markers are available for use, and they vary in their reliability and ease of assessment [23]. Generally, deep neural networks are used to perform biometric identification given the large, complex form of their data. In this system, a neural network to perform the task of biometric identification needs to be developed with constraints outside of that found in existing literature resulting from the use of blockchain.

1.4 Technical Innovations

Success in the aim of this project is primarily dependent on the achievement of three technical innovations. The first of these is to develop a biometric identification model that can produce a form of data compact enough for storage on blockchain but reliable enough when used for identification to be fit for purpose in the system. This is a unique challenge, as generally restrictions in neural network design come in the size and performance of the model itself, not the output it produces. Reduction by the extent required for compatibility with blockchain is around a factor of 1000 compared to ordinary applications.

The second major innovation comes in the development of a certification system that can be stored on blockchain and remain compliant to UK GDPR restrictions. Additionally, it must do this while being able to prevent acts of shared identity fraud. Given the fact that GDPR was not designed with blockchain in mind [24], measures that go far to preserve user anonymity are required.

The final innovation comes in implementing the blockchain-based solution in a way that integrates seamlessly with web and app use cases. This means it must be free, fast, and easy to use for day-to-day activities. Given the costs and delays that can be experienced with blockchain projects, this is a major challenge to overcome. However, achieving it gives users a system where personal data are only stored on their devices, and verifications are accessible, reliable, and transparent.

Chapter 2

Background and Related Work

The area of digital identity management is well studied and the potential challenges that such systems may face is well documented. As such, there is a wealth of information available from previous reports and attempts at providing solutions that were used to inform the development of the system presented in this report. Additionally, blockchain has been a topic of many studies, particularly over the last few years as its popularity amongst academics and the general public alike has rocketed. This information has been equally useful in informing the design of the blockchain-based digital identity system.

The following chapter presents findings that proved instrumental in the design process of the system. Sources of particular insight and significance have been addressed and cited. Where possible, studies of high importance to the field they cover which are highly referenced and peer-reviewed have been selected to acquire the most accurate and accepted information and points of view available. As blockchain is an area of rapid development, dates of works have been noted appropriately to account for any developments that have emerged in recent years.

2.1 Blockchain

Blockchain has undergone a fantastic rate of development and a phenomenal rise in popularity since its inception with the 2008 publishing of the Bitcoin whitepaper by 'Satoshi Nakamoto' [5]. In that time, a large number of other blockchain protocols have been developed and a variety of use cases implemented [25]. Any application that wishes to use blockchain must either establish its own protocol or rely on an existing layer 1 protocol, such as Bitcoin or Ethereum, which are able to validate transactions without the need for another network. In this project, a layer 1 protocol will be selected to build on that meets the needs of the system.

A blockchain is fundamentally a distributed database. It is made up of nodes that maintain the information and state of the network. It records transactions between parties and relies on validation of those transactions through consensus between the nodes of the network. Information is grouped into blocks of a set size. When a block is filled, it is added to the chain of existing blocks that details all past events and information on the blockchain [5]. The design of Bitcoin's blockchain

brought about new possibilities. By relying on a consensus algorithm called 'proof-of-work', it was a complete redesign of conventional consensus systems. Previously, only trusted, identifiable parties would be allowed to participate in such setups in case an attack was attempted. In the new system, anyone can be involved in the consensus process. Weight in the consensus vote is decided by computational power, making the system fair and protecting the network from sybil attacks by creating a barrier to entry and influence [26]. This leaves the only major vulnerability of the proof-of-work approach a '51%' attack [27]. This is when a single node or party is able to get control of over half of the computational power of the network, allowing them to create and validate blocks at their leisure. Therefore, a successful blockchain project that has achieved a high degree of decentralisation is very robust.

Other Consensus Algorithms

Since Bitcoin, alternative consensus mechanisms have been suggested [28]. Each of these aims to solve different problems including environmental cost, scalability, and speed. The most prominent of these is proof-of-stake which moves the burden of the algorithm from computational strength to the amount of cryptocurrency held by a node under the belief that a node is unlikely to attack a network it holds a large share in [29]. This is deemed a more environmentally friendly option as it does not require the expensive procedures used in a proof-of-work algorithm. Alternatively, proof-of-history [30] relies on a sequence of computation to verify the passage of time between two events. It uses hash functions such that the output cannot be predicted without following the computation through each step in the process. It is the fastest of the consensus algorithms and can give rapid verification. In systems that wish to take on a more old-fashioned approach, proof-of-authority can be used [31]. This is a form of consensus where only pre-approved individuals can act as signers and authenticate transactions. It goes against the traditional ideology of blockchain and the approach that 'Satoshi Nakamoto' pioneered as the network can no longer support any individual as a participant. It does, however, still achieve decentralisation and establishes a protocol for agreement between nodes. Thanks to the implicit trust in the authorised signers, this form of consensus requires no computational effort, staking, or other efforts by the signers to prevent attacks on the network [32].

Permissioned and Permissionless Networks

Blockchains like the Bitcoin and Ethereum mainnets are permissionless. This means that anyone can join them, either as a user or as a node, and the data published on them is open to the public. This form of blockchain network makes up the vast majority of existing projects as they fall in line with the beliefs central to the technology's inception. They are completely indiscriminate and are owned, run, and maintained by their users.

In a permissioned blockchain, only authorized parties can join or mine [33]. These are of particular use in cases where confidential information is being handled or limitations on the participators of the distributed application (DAPP) must be

enforced. Therefore, a system hoping to manage identities may be built on a permissioned blockchain. If not, care must be taken in ensuring no private data is stored in the network and access is carefully controlled to prevent modification by bad actors. A permissioned blockchain can use an established protocol such as Ethereum, but the blockchain itself will be separate from the primary blockchain associated with that project. A permissioned blockchain has its own financial ecosystem. Transaction costs and mining incentives can be set at will, and the difficulty of mining can be set based on the potential risk of a 51% attack [34].

2.1.1 Comparative Analysis of Layer 1 Protocols

Table 2.1: Comparative analysis of top layer 1 protocols.

Protocol	Transactions per Second	Deploy Smart Contracts	Permissioned Blochain Support	Consensus Algorithm	EVM Compatible	References
Bitcoin	7	∼	✗	PoW	✗	[35]
Ethereum	15	✓	✓	PoW	✓	[35]
Binance Smart Chain	160	✓	✗	PoS and PoW	✓	[36]
Solana	65,000	✓	✗	PoH	∼	[35]
Avalanche	4,500	✓	✓	PoS	✓	[35]
Cardano	250	✓	✓	PoS	✗	[35]
Polkadot	65,000	✓	✓	PoS and PoW	✓	[35]
Algorand	1,200	✓	✗	PoS	✗	[37]
Terra	10,000	✓	✗	PoS	✗	[38]
NEAR Protocol	100,000	✓	✓	PoS	✗	[39]
Cosmos	10,000	✓	✓	PoS	✗	[35]

A comparative analysis of the layer 1 protocols available for use in this project is presented in Table 2.1. For selection in this project, each of the criteria must be carefully considered for the identity management use case. Transaction speeds only apply to interactions that involve the modification of data on a smart contract as these require verification through the consensus mechanism. Interactions that simply read data from the smart contract will consistently run quickly, primarily dependent on internet connection. Additionally, these interactions have no gas fee as they are a communication between the user and a single node of the blockchain [40]. It should be noted that figures given in Table 2.1 are derived from each protocol's main chain. When deployed on a permissioned blockchain, parameters can be modified and node organisation adjusted to alter the performance with attention paid to the security of the network in the context of this specific application.

Alongside the information presented in Table 2.1, the support for each of these protocols and the potential for it to work with a variety of different software solutions should be considered to make the identity management system as extensible and practical as possible. Ethereum is by far the most broadly accepted blockchain for building DAPPs [41]. As can be seen in Table 2.1, numerous other protocols are compatible with the Ethereum Virtual Machine (EVM) so that they can run the same smart contracts and work with the same third party libraries [42].

2.1.2 Blockchain Storage Limitations

Blockchain is not designed to store large quantities of data. Since an immutable record of all transactions is kept, and every node on the blockchain has a copy of this, if large quantities of data were stored this would see the size of the blockchain swell to petabytes of data in little time and grind the network to a halt. This limitation in the ability to store data had to be taken into account while designing the system. The certifications of users' identities that will be stored on the network must be compact yet effective.

In Ethereum's smart contract programming language, Solidity [43], there is an upper limit to the size of data types provided. This is 32 bytes [4]. It is possible to split data across multiple containers in order to bypass this restriction, though this is bad practice and will incur speed penalties and larger gas fees for data writing and modification.

2.1.3 Blockchain Topologies

Blockchain topology refers to the way in which a blockchain's nodes are organised, run, and incentivised [44]. It can have large implications on the performance, security, and practicality of a network. Multiple options for the topology of a blockchain network exist and primarily vary on the level of decentralisation they achieve. The options available for use in a digital identity system must be considered with both the technical strengths and practical costs in mind.

The following points outline the main strengths and weaknesses of each blockchain topology option available for the system:

- **Public Mainnet:** The construction of the system used to record the state of certifications of users within the network can be made compatible with a blockchain mainnet. Access restrictions can be put in place such that only authorised individuals can make changes to the state of the system. Additionally, private information can be withheld from the blockchain, meaning that in the case that information is accessed, with or without permission, it will not compromise the security of the user [45].

  Strengths:

  - This is the most decentralised solution for blockchain hosting thanks to the existing size of layer 1 protocols [46].
  - It comes with tried and tested network performance, reliability, and security.
  - This requires minimal organisational overhead as this is reliant on existing functionality.

  Weaknesses:

  - The system would be restricted to the speeds achievable by the mainnet for data-modifying transactions.
  - The system would be vulnerable to network crashes, slowdowns, and attacks that occur on the mainnet and are not related to this project [47].
  - To avoid account management by users who may be non-technical or uninterested, and to also avoid the distribution of value tokens to users which may be abused by bad actors, all data-modifying transactions must be completed by a centralised governing authority.

- **Permissioned Blockchain Hosted by a Node Provider:** Blockchain node providers are services that can be used by enterprises to host private blockchains [48]. Each make their own claims on the privacy, decentralisation, and reliability of their services. For a set cost, a service such as this could be used to host the blockchain for the identity management system.

  Strengths:

  - This solution has the potential to have good decentralisation.
  - It comes with low organisational overheads as these services are already established. Additionally, nodes do not need to be convinced to join the network before it becomes usable.
  - Removal of a value system in the chain makes data-modifying transactions far simpler since there are no gas fees [49].

  Weaknesses:

  - Claims of decentralisation and security made by the provider may be untrue or misleading [50].

  - The governance of nodes, however they may be arranged, by a single entity (the provider) reduces the decentralisation of the system.
  - The system is reliant on the availability of the network which is completely out of the control of system administrators.

- **Permissioned Blockchain Hosted by a Custom Network:** The final option for the blockchain is to create a purpose-built permissioned blockchain. Since this does not rely on any existing network, nodes would need to be encouraged to join the network in some way. This is a challenging problem alone, as it requires some form of compensation for nodes which will ultimately become a cost for users. Additionally, in its early stages this network would be very vulnerable to 51% attacks if a proof-of-work consensus scheme is used as the network will not be backed by very much computing power. Smaller custom networks like this tend to take on the proof-of-authority scheme defined in Ethereum's standards to avoid this issue [32]. In this setup, only authorised signers can mint blocks. New signers can be introduced in a voting protocol. This solution risks either being too centralised if signers are added tentatively, or too vulnerable if signers are added without vetting.

  Strengths:

  - Removal of a value system in the chain makes data-modifying transactions far simpler since there are no gas fees [49].
  - All reliance on third parties and coupling to other projects is removed.
  - Chain parameters can be selected at will meaning that high performance is achievable [34].

  Weaknesses:

  - Node organisation and compensation is challenging to solve and could lead to poor decentralisation or security.
  - It comes with a high organisation overhead as an entire the network would need to be designed and implemented.

2.1.4 Blockchain for Digital Identity

The core strengths of blockchain are privacy, security, anonymity, decentralization, and immutability [25]. A systematic literature review by Dr Jaoude of potential blockchain applications highlights that any area in which these characteristics are valuable could potentially benefit from a blockchain implementation [51]. Hence, areas such as digital identity where all of these factors could be of value stand to benefit greatly.

The consensus-based design of a distributed ledger makes it near impossible to establish fraudulent identities. Unless an attacker was able to overwhelm the network by controlling over half of the consensus votes (a 51% attack) the ledger

cannot be changed [27], and given the infeasibility of such an attack, it forms one of the most fraud-resistant options available.

Blockchain also offers a decentralised environment where there is no trusted third party [52]. There is therefore no trust required in an entity to protect the data or act in the best interest of users. In the context of an identity system, this is of the utmost importance, and is crucial in empowering users with the control of their own data, as highlighted in the 'Laws of identity' [12]. Time after time, the technology industry has faced controversies over the handling of data, whether it be immoral, such as the Facebook–Cambridge Analytica data scandal where personal data were collected without consent for political advertising [53], or negligent, such as the 2019 Instagram data leak due to an unprotected server full of personal information [54]. The decentralisation of a blockchain-based design bypasses these issues and puts users before corporate interest.

2.2 Legislation and Digital Identity

Any digital identity management solution must protect the valuable private data provided by users for their identity. This is enshrined in the UK's information rights, given in the UK GDPR [2]. Over the course of this section, a number of sources referring to both the EU GDPR and UK GDPR are cited. These are almost word-for-word copies of each other with the UK GDPR being established in UK law upon Brexit. The only exception to this is the listed enforcing bodies of the regulations [55].

The system developed in this project will conform to the GDPR, as required by UK law. An important area of this to note is the 'right to erasure'. This protects the right of any user to request their personal data be deleted at any point in time. Within GDPR, personal data are defined as 'any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person' [2]. Blockchain is, by definition, immutable. Therefore, any data stored on the blockchain, encrypted or not, cannot be deleted. It is impossible for a system relying on blockchain for personal data storage to conform to UK GDPR, so this approach cannot be taken. Instead, an approach where data that can validate personal details reported by an individual are stored on the blockchain can be used, where this data does not fall within the given definition of personal.

GDPR Data Classifications

All data is classified as one of two categories in GDPR. They are either personal data or anonymised data. Converse to the definition provided for personal data, anonymised data are data that have been made unidentifiable through any reasonable means, effectively making them impossible to reverse [56]. Anonymised data

are not subject to GDPR regulation and the right to erasure does not apply to them. Pseudoanonymised data are data where the individual's identity has been somehow obfuscated, yet they remain identifiable through some means of reversal or by combination of multiple different data [57]. Pseudoanonymised data are classified as a form of personal data in GDPR, so they are subject to the right to erasure and not suitable for storage on blockchain.

The definitions of personal data and anonymised data provided in GDPR are intentionally vague. Data's classification relies heavily on the context they exists in [58]. There is no simple rule to follow that ensures GDPR compliance when it comes to the requirement for conformity. Instead, it is up to the data holder to carefully consider the possibility for identification of individuals from the data and make their case for categorisation [59]. The vagueness of the legislation has led to multiple interpretations of the definitions being taken. The most egregious example of this comes from the Article 29 Working Party itself, the commission responsible for designing the regulations. In a 2007 opinion publication on 'the concept of personal data' [60], a clear statement was issued on the difference between anonymised and pseudoanonymised data, requiring only 'appropriate technical measures' to be put in place to prevent reidentification for data to be classified as anonymised. It went so far as to say that 'disguising identities can also be done in a way that no reidentification is possible, e.g. by one-way cryptography, which creates in general anonymised data'. This makes hashing as a form of anonymisation explicitly permissible. However, a later 2014 opinion published specifically 'on anonymisation techniques' [61] had far harsher criteria. It specifies that anonymised data 'would be robust against re-identification performed by the most likely and reasonable means the data controller and any third party may employ'. While this sounds familiar to the definition provided in the 2007 opinion, it explicitly states that hashed and encrypted data does not meet the requirements of anonymisation due to the possibility of dictionary attacks and key cracking. However, it is noted that hashing can be permissible if the input is so unpredictable that identification through brute force attacks becomes utterly infeasible. It also highlights that 'aggregating data into group statistics' is a strong way for organisations to ensure that the data they use is properly anonymised.

Working with Blockchain and GDPR

In more recent years, the challenge of squaring GDPR with the immutability of blockchain has been studied in more detail. The exact requirements for anonymisation remain as foggy as ever, with a 2018 Delphi Study [62] looking into privacy protection on blockchain with GDPR concluding that 'hashing could be a valid solution to store personal data on the blockchain'. Meanwhile many other studies reason that GDPR is unavoidably incompatible with blockchain and hashing of personal attributes is not suitable for anonymisation [63, 64, 65]. Based on these sources, the overwhelming majority of both direct communication from governing bodies, as well as studies into the legality of different options, favour more strictly defined rules on anonymity, making construction of an identity management solution on the blockchain highly challenging.

While GDPR was obviously not designed with blockchain in mind, its popularity

in recent years has spurred responses from major data protection bodies, including CNIL, France's data authority [66]. They note the impossibility of data deletion on blockchain, and cite the disposal of encryption keys or smart contract access revocation as suitable alternatives in the circumstances. They also note that in many cases, blockchain serves to increase privacy and transparency for its users, making the conflict with GDPR feel cumbersome, outdated, and counterproductive.

Traversing GDPR when working with blockchain technology is a legal minefield. Not only are the criteria hard to meet, they are unclear. This leaves the best approach for remaining compliant as minimising the possibility for reidentification in every feasible way. The literature consulted in this section has highlighted two allowable ways of anonymising personal data. Firstly, data can be grouped in such a way that no individual can be identified. Secondly, one-way cryptography techniques such as hashing can be used given the input is unpredictable, long, and utterly robust to dictionary attacks.

Government Action on Digital Identity

The Government has already expressed interest in implementing a digital identity scheme. Early digital identity testing is ongoing for use in right to work, right to rent, and DBS checks [67]. This is laying the foundation on which a digital identity management system could be built. Core to the work so far is the UK digital identity and attributes trust framework [1], which prototypes the rules that must be adhered to for parties involved in the world of digital identity, including identity service providers, the category this project would fall into. Given this information, all development was done in line with the current guidance to ensure its feasibility. The guidance remains relatively preliminary as the document is still in an Alpha release. The sections relevant to identity service providers currently dictate that they must adhere to GPG 45 [9] and prove that their system is safe, secure, and reliable. These requirements should be covered over the body of this report. The Government has specified that they will be leaving the development of digital identity systems to private parties, meaning that concepts such as the one being explored in this project have potential for real world adoption.

2.3 Identifying Information for Digital Systems

For an identity management system to function, there must be some quantity of information that can uniquely identify users. In a digital context, this is required to prevent multiple accounts from being set up using the same identity. Various options exist that could be used as this identifying information. It is vital that any information selected can be verified. For instance, if name and address were selected, the authentication process would require the submission of an ID along with an official document with the name and address.

Options available for the unique identification of users and their strengths and weaknesses are summarised below:

- **Name and Address**: Typically, name and address are used to authenticate identity in lower security settings where the risk or consequences of fraud are lower. This is because official documentation in the form of e-letters are easy to forge or modify and home addresses can change on a regular basis [15]. In the context of a secure identity management system that could give total control over aspects of an individual's life, the risk and consequences of fraud are high. Therefore, this solution does not pose sufficient security to be used.

- **Document Reference Numbers**: Almost all forms of identity documentation have a unique reference number. Initially, this may seem a good option to uniquely identify users of an identity management system. However, this solution suffers from a major shortfall. These reference numbers are associated with the document and not the individual. Therefore, different numbers and numbering systems are used across different forms of ID [68, 69]. In order for this approach to be used within an identity management solution, users would have to be limited to a single form of ID, such as passports, for the authentication process. This seriously impacts the accessibility of the system as not all people will have access to this document. The UK has no national identity card and no requirement to hold a particular form of ID [70].

- **National Insurance Numbers**: A stronger solution is to use national insurance numbers. Though not designed to be used for identification, it has widely been adopted in the financial space for that very purpose [71]. Originally, national insurance numbers were simply supposed to be used in the administration of national insurance, the UK's social security scheme [72]. Its accidental adoption as a unique identifier is evident from the fact that a national insurance card cannot be taken as proof of identity [73]. The reason it has taken on the utility that it has comes from the fact that the UK has no national identity card or national identity number [70].

  The UK has been opposed to any compulsory identification credentials in spite of widespread adoption across Europe. Attempts to bring forms of national identity in, such as the 2006 Identity Cards Act [70], were met with protests and criticism [74]. At the core of the debate lie issues of privacy, security, and human rights. Those against the initiatives are not comfortable with the potential requirement of citizens to carry an ID card and to have to identify themselves to figures of authority. Others do not want to see a centralised database with all sensitive, personal information stored at the hands of the Government [75].

  Since all UK citizens over the age of 16 are issued with a national insurance number, it has become a convenient alternative in absence of a proper solution. However, it is not available to all bodies. The Government tightly controls the use of national insurance numbers due to their relationship with finances and potential for fraud. It is therefore advised that you do not share your national insurance number with any entity except those from a list of government approved financial institutions [76]. Furthermore, only the Government has the capacity to verify a user-reported national insurance number as being correct,

and they will only do this for the aforementioned financial institutions [71]. It is therefore not available for use in the vast majority of situations and has led to the NHS, DVLA, and police force developing their own systems [77, 69, 78].

Therefore, national insurance numbers cannot be used as identifying information in the digital identity management solution. Furthermore, the fact it was not designed for widespread identification purposes make it a less than optimal solution with the availability of technology for biometric solutions.

- **Biometric Data**: Biometric data forms the strongest way to identify an individual as it relies on 'something you are' as opposed to 'something you have' [22]. Unlike other credentials, it cannot be passed onto another individual and is very hard to change as it requires physical alteration of your body. Furthermore, as opposed to other forms of identifying information, it does not rely on an existing centralised body to issue the data. This furthers the decentralisation of the system's design as it uses intrinsic facts about an individual to make the identification rather than fabrications. Biometric data are typically seen as the gold standard of identification, forming the highest level of confidence in GPG 45 [9]. In the spirit of using the 'least identifying information', biometric data are also strong. While offering great confidence in confirming an individual's identity, it is meaningless on its own and cannot be used to compromise the privacy of an individual in the same way an address or similar attribute could.

2.4 Biometric Identification

Biometric identification is the recognition of individuals based on biometric markers. These are physical qualities of the human body that exhibit particular uniqueness in the greater population and have some measurable feature [22]. Their use has become widespread from fingerprint detection in mobile phones [79] to facial recognition in security cameras [80]. They are favoured for these and many more use cases due to the fact that they are not easily modified and are intrinsic properties of every human [81]. As a fantastic tool for identity verification, they have been and are being introduced into identity management systems from nationwide passport programs [82] to government-led digital identity trials [83].

Comparative Analysis of Biometric Markers

With a wide selection of biometric markers available, their use in any identity management system should be carefully decided based on the accessibility, integrity, and speed of their implementation. In Table 2.2, a comparative analysis of available biometric markers with respect to these metrics is presented. Any marker that can achieve data collection with a smartphone or similarly common device has been deemed accessible. The reliability and speed of validation has been scored based on the methods used to assess each of the markers for practical identification purposes.

Table 2.2: Comparative analysis of biometric markers.

Biometric Marker	Accessibility of Measurement	Reliability of Verification	Speed of Verification	References
DNA	**Low** Easy to collect, hard to assess	**Very High**	**Very Low** Specialist equipment and lab required	[84]
Ear Shape	**Medium** Requires little specialty	**Medium/High** Dependent on extent of measurement	**High** No specialist process required	[85]
Iris	**Medium** Requires suitable camera	**Very High** Highly unique attribute	**Very High** On data collection, automated verification can be done rapidly	[86]
Retina	**Low** Requires specialist camera	**Very High** Highly unique attribute	**Very High** On data collection, automated verification can be done rapidly	[87]
Scleral Vein	**Very Low** Requires highly specialised equipment	**Very High** Highly unique attribute	**Very High** On data collection, automated verification can be done rapidly	[88]
Face	**Very High** No specialised equipment	**Medium/Low** Relatively unique attribute	**Very High** On data collection, automated verification can be done rapidly	[89]
Finger Geometry	**Low** Needs specialised equipment and supervision	**Medium** Relatively unique attribute	**Medium** Not a rapid or automated process	[90]
Fingerprint	**High** Available on many smartphones	**Very high** Highly unique attribute	**Very High** On data collection, automated verification can be done rapidly	[91]
Walking Gait	**Very Low** Needs specialised, bulky equipment and supervision	**Medium/High** Dependent on extent of measurement	**Low** Requires detailed analysis	[92]
Hand Geometry	**Low** Needs specialised equipment and supervision	**Medium** Relatively unique attribute	**Medium** Not a rapid or automated process	[93]
Heartbeat	**Medium** Relatively widespread equipment in smart watches and phones	**Medium** Difficult to achieve precision based on data	**High** On data collection, automated verification can be done rapidly	[94]
Odour	**Very Low** Very specialised equipment	**High** With proper technique, uniqueness is high	**Very Low** Requires special equipment and lab	[95]
Vascular	**Very Low** Very specialised equipment	**Very High** Highly unique attribute	**Very Low** Requires special equipment and lab	[96]
Voice	**Very High** No specialised equipment	**Medium/Low** Not precise for large populations	**Very High** On data collection, automated verification can be done rapidly	[97]

2.4.1 Neural Networks for Biometric Identification

Many of the biometric markers identified are made up of huge quantities of information. They are not a single measurement, but instead the combination of a number of variables that together are unique to each person. For instance, fingerprints would be nowhere near as uniquely identifying if they simply measured the spacing between indentations on the finger. Instead, the entire pattern must be taken into account.

The complexity of these markers makes them difficult to compare by conventional means. Although a face may not appear identical in different photos, and the photo data itself will always vary, the system used to make biometric verifications must be able to process this and still make accurate classifications.

Based on the nature of biometric verification, neural networks are often used to provide fast, reliable results. They have been found to greatly outperform alternative algorithms used for facial recognition [98], fingerprints [99], or iris scans [100]. Neural networks are able to extract information from the measurements and detect highly complex relationships in the data. For biometric information to be used in the digital identity management solution, a neural network capable of carrying out biometric marker verification will be required.

Siamese Neural Networks for One-Shot Image Recognition

An example of a neural network that can be used for biometric recognition is given in Gregory Koch et al's paper, 'Siamese Neural Networks for One-shot Image Recognition' [3]. This details a model architecture that can take two inputs of facial images and returns either a one or a zero to signify a match or not. It has received over 2500 citations since it was published in 2015 highlighting its significance in the field and it forms a strong yet compact image recognition model.

This form of architecture is particularly interesting in its ability to make classifications on unlearned images. The model is trained by feeding it a set of matching and not matching inputs. It can then be applied to objects or instances it has never encountered and remain reliable.

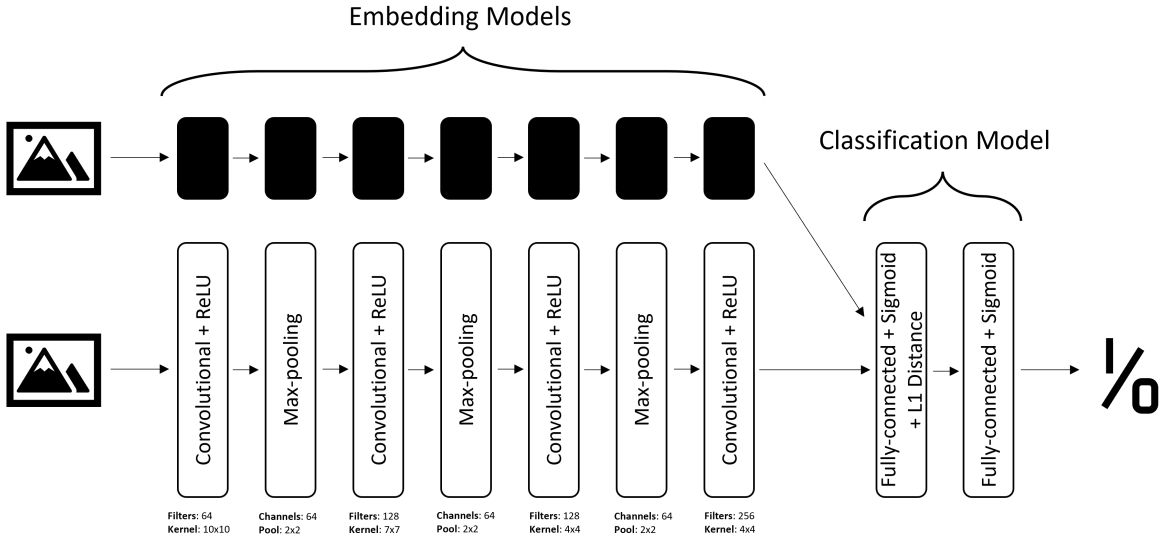

Figure 2.1: The model architecture used in the neural network proposed in Gregory Koch et al's paper, 'Siamese Neural Networks for One-shot Image Recognition' [3]

The architecture used for this model is presented in Fig. 2.1. It consists of two identical embedding models and a classification model. An embedding model is a form of neural network that seeks to extract features from an input [101]. As

opposed to traditional approaches where these features may be hand selected, the model is allowed to identify the aspects of the data that lead to the greatest degree of identifiability. This makes it particularly well suited to deal with the complex, large data present in biometric markers.

The embedding model is made up of convolutional layers. These are often used for computer vision tasks as they are particularly good at extracting dense information from image data. The exact choice of the hyperparameters used in the model presented in Fig. 2.1 are the result of a random hyperparameter search conducted by Gregory Koch et al to cover all reasonable values for layer kernel size, pooling size, channel number, and filter number.

Figure 2.2: Examples from the Omniglot [102] data used for training in Gregory Koch et al's paper, 'Siamese Neural Networks for One-shot Image Recognition' [3]. The data contains symbols from many different languages.

In Gregory Koch et al's paper [3], the model was used to identify symbols from the Omniglot [102] dataset as the same or not. Examples of the images used for classification are shown in Fig. 2.2. While this task may seem easier to achieve and less intensive than biometric classification, the model can easily be applied to new purposes by simply selecting appropriate training data. In its original application, it was able to achieve an accuracy of 93.42% against a human baseline of 95.5%.

Neural Networks on Blockchain

Any personal data being handled in an identity management system is at risk of privacy or security infringement. One of the major benefits of integrating blockchain is that it provides transparency ensuring that these infringements do not occur. Therefore, in an ideal world, delicate processing would be handled on the blockchain.

It may therefore be desirable to process the biometric information for identity verification on the blockchain. This would require the algorithm responsible for this job to be hosted on the network. While the concept of running neural networks on blockchain is not new and has been explored in existing research [103], it is not feasible with the current technology.

The main barriers to the possibility of neural networks running on the blockchain are firstly the computational expense of the process. For blockchain transactions, every node of the network must run the full computation. This would have an immense power, time, and money requirement and as such is completely impractical.

A system utilising some form of sharding may be able to deal with the challenging computation by sharing load across nodes, but this is not a feature available at the moment [104].

The second major barrier is the nature of neural networks and the way in which they are run. Typically, GPU clusters are used for their high-performance parallel computation ability [105]. These tend to offer non-deterministic results due to the possibility of race conditions [106]. While inconsequential to the design of most neural networks, this is simply not compatible with blockchain technology which expects deterministic outputs to smart contracts in order to fulfill the consensus that the entire technology is based on. This would be a very challenging hurdle to overcome given how essential consensus is to the nature of blockchain systems.

2.5 Digital Identity Management in Literature

A number of solutions for digital identity have been proposed and developed both in literature and commercially. These, along with their main features and shortcomings are highlighted in Table 2.3 and discussed throughout this section.

Table 2.3: Evaluation of the features and shortfalls of proposed digital ID solutions

Project	P2P Verification	Full Identity Information Certification	Decentralised Data Storage	Network Architecture	GDPR Compliant Solution
Namecoin	✗	✗	✓	Public	Potentially
I/O Digital	✓	✓	✓	Public	✗
Sovrin	✓	✓	✓	Permissioned	✗
uPort	✗	✗	✓	Public	✗
Bitnation	✓	✓	✓	Public	✗
UniquID	✓	✗	✓	Public	✗
Jolocom	✓	✓	✓	Public	✗
Cryptid	✗	✗	✓	Public	✓
Company Identity Management Proposals	✗	✗	✓	N/A	Potentially
Dr Mudliar's National Identity Proposal	✓	✓	✓	N/A	✗
e-Residency	✓	✓	✗	N/A	✓
1account	✓	✓	✗	N/A	✓
PASS 5 Standard	✓	✓	✗	N/A	✓

The potential for blockchain to be used for identity management systems has been noted before this point. One of the earliest adoptions of blockchain was in

2010's Namecoin [107]. This was a name registration database that allows users to claim a unique identifier in a first-to-file system. The Ethereum whitepaper [108] cites this project and also identifies the potential for DAPPs to function as identification systems thanks to the immutability and distributed nature of the technology.

Many projects have taken measures that extend beyond name ownership in the effort to digitise identity. I/O Digital is a Bitcoin side-chain that claims to be the first decentralised identity service [109]. Built on a similar premise to Namecoin, it allows users to attach their personal details to a specific bitcoin address, giving users the ability to use blockchain cryptography for purposes outside of the network for the first time. It lacks any real oversight in this process, meaning that credentials are self-reported. This makes it useless for any regulated processes, but it does support P2P verification of these self-reported credentials.

Sovrin is a blockchain-based attempt at forming an internet identity layer. In its whitepaper, many of the core concepts expressed in valuing blockchain for the empowerment of users' autonomy over data are expressed [19]. The system is based on a permissioned blockchain where signers are selected in a distributed, vote-based trust system. Users are issued accounts that act as their unique identity tokens. Lists of authenticated claims such as name or date of birth are then stored on the blockchain in an encrypted form so that they are only accessible to the user they belong to. When a user needs to prove one of their claims, they can choose to either reveal it directly to the verifier, or they can use zero-knowledge proofs to prove a given property, such as knowledge of an account number, without revealing it. This solution suffers from a major shortfall. Personal data are stored directly on the blockchain, making this option unusable in the UK and Europe since it does not conform to GDPR standards [2]. Other projects, such as uPort, a slightly less comprehensive account management solution, suffer from similar problems. In this scenario, user login details are stored on the blockchain [20]. Once again, this fails to meet GDPR as email addresses count as identifying information since they are reasonably traceable and associated with a single user.

Alongside internet-focused solutions, some projects seek to make more radical changes to the ways in which people manage identity. Bitnation is a decentralised pseudo-nation of voluntary citizens [110]. It is founded on the idea of modern nationality based on ideology as opposed to geography. The organisation supports storage of records and identity on Ethereum, services which would usually be the responsibility of an individual's country of residence. It offers its users a new decentralised structure where they are in control of their own data, and no centralised body is able to make decisions over their life. Instead, the system is entirely peer-to-peer, with its members in governance over the project. It has been broadly cited as a powerful tool for under-represented groups that lack power and agency in their nations [111] and some sources have even claimed that it shows how blockchain can be used to help solve the migrant crisis [112].

In one of the most practical blockchain identity projects to date, Jolocom provides a decentralised, self-sovereign 'wallet' for the storage of credentials and documents [113]. It can be used for a wide range of purposes including identity, tickets, and corporate management. As a fully decentralised service, there is no central over-

sight. However, documents are issued by a recorded party. This might be the user or some third party. A verifier can view the party who issued the attribute, and decide whether or not they will accept it based on the trust they have in them and the job it is being used for.

Blockchain identity management can have use cases besides individual identity ownership proofs. UniqueID is an open-source project aiming to provide decentralised identity and access management [114]. It is more of a general protocol rather than a specific use case, and provides minimal trust networking. They claim the solution is more efficient than traditional protocols for authenticating attributes and point to potential use cases in cloud services, IoT, and machine-to-machine communication. Of course, the protocol can be used to store and prove any information, though the management of this directly through blockchain makes the protocol incompatible with GDPR.

Innovation also exists in the way identification on the blockchain itself works. Cryptid is an identity layer for the Solana blockchain [115]. It seeks to remove ownership of blockchain addresses from single private keys, and instead offers a service that allows users to effortlessly manage multiple wallets and remove the risk of a single private key being compromised. It demonstrates an alternative method of blockchain identity management and verification, humanising the process to make it more user-friendly and secure.

Company identity management refers to policies ensuring that only validated individuals can access given resources. This area has also been highlighted as having potential for blockchain implementation in reviews of potential applications of the technology dating back to 2018 [116]. In these cases, the reliable access of blockchain and the ability to shift sensitive data from servers to user devices was noted. This has led to major companies developing private blockchains for internal usage, whether it be for identity management or other use cases. One such example is Quorum, developed by JP Morgan, a private blockchain protocol built on top of Ethereum for commercial use [117]. However, it should be pointed out that some literature on the topic of company identity management has acknowledged the difficulty in changing user identity in a blockchain-based system [118]. This is a difficulty that extends to broader attempts at implementation.

A digital identity system will always need some amount of government support to be adopted in situations protected by legislation, such as the purchase of controlled goods. Some proposals in existing literature go as far as to claim that all national identification records should be stored in a blockchain-supported system. One such proposal by Dr Mudliar in 2018 pointed to the difficulty of record alteration and the transparency of such a system as the primary benefits [119]. In this paper, it was clear how the ability to program the blockchain into certain behaviour through smart contracts that are visible to all would create a system in which every party was fully aware of how their data was being used. However, this proposal was designed with India in mind, and does not meet GDPR regulation. Furthermore, the storage of data on the blockchain leaves it exposed to hackers, even if it is encrypted. A system that stores no form of recoverable data on the blockchain would be preferred. Another paper by Dr Chalaemwongwan [120] highlighted the power

of blockchain to centralise national identity. While this may seem paradoxical, the paper pointed out the many different forms of identification and identifying information used for different government bodies and other institutions. The ability of a blockchain-based digital identity solution to create immutable identifiers for each civilian, and the ability to program smart contracts to expose only the necessary information to parties as it is required, seems a strong solution to this issue.

Digitisation of identity management opens opportunities to unravel traditional limitations of borders. This has been recognised by the government of Estonia in their e-Residency program [121]. This is a scheme to allow anyone from any country to apply for a digital residency and identity associated with Estonia, allowing them to access government services and attain government-affiliated documentation. Its greatest benefit is perhaps for those who have been pushed aside or forced out in their own countries, though the service is indiscriminate and open to everybody. While not decentralised, the idea of a cross-border identity service to help the most vulnerable in society shows the ethical good achievable with a digital identity scheme.

Other non-blockchain-based proposals for digital identity have also been developed and tested. A government-sponsored trial by 1account is currently running tests in a nightclub to phase digital IDs into the UK for nightlife [122]. Door staff have been given special training on using the service. However, 1account relies on centralised storage of user data. It, therefore, does not treat their data with the strict protocols required of an ethical digital ID implementation. Moreover, the service uses a relatively rudimentary verification system. A security code is displayed on screen that updates every minute [123]. While this does prevent the vast majority fraud attempts, it leaves unnecessary gaps open for exploitation when rolled out on a nationwide scale. This process also provides a cumbersome UX, requiring door staff to manually enter the code. If a code change occurred halfway through a verification, the process would need to be restarted.

Other parties are working to establish what digital identity may look like. Proof of Age Standards Scheme (PASS) is a Government affiliated body who issues accreditations to ID card manufacturers who meet a set of standards [124], reassuring the establishments they are used in that they can be legally accepted. They have expressed interest in the development of a digital proof-of-age by publishing a consultation with the public into the development of such a system, which collected insights from a range of participants on aspects they believed important to the future of digital proof-of-age in the UK [125]. In May 2022, PASS published a PASS 5 alpha, their Requirements for Digital Presentation of Proof-of-Age [126]. This is a preliminary attempt at laying out what would be required for a digital proof-of-age system. It covers areas including the presentation of the app and the security practices required. While conformity to these standards will not be required by law, it does give a good indication of what the Government may formulate. It is noted that the security requirements laid out in the standard do not go far to protect user data and rely on typical practices with confidential data. For the most part, this does not seem unreasonable. However, the frequent history of major data leaks from entities with similar practices highlights the shortcomings of this approach [127]. While re-

liance on this is somewhat difficult to remove completely, other approaches should be investigated when data of such value and potential risk is at stake.

As shown in Table 2.3, no current proposals are able to satisfy the requirements of decentralisation and GDPR compliance while offering reliable P2P identity verification. There is good reason for this. Preventing multiple accounts being established from the same identity when there is no centralised data storage and the data itself cannot be placed on the blockchain is very challenging.

2.6 Background Reading Conclusions

The background covered through this chapter has led to several key conclusions that were taken forward to the development of the identity management system. It is clear that blockchain provides a secure, private, and decentralised platform for data storage that is perfectly suited to identity management [25]. A variety of different protocols and topologies are available to meet the exact needs of the system. The degree to which the blockchain is decentralised is seen to be of particular importance as this is the source of the technology's security and user control [5]. It is therefore a crucial factor in the decision-making process of this project. The impact of the decisions related to the blockchain also affects the integrity and portability of the solution [44]. For the digital identity management system to be treated as a reputable source of information and extensible to a broad set of applications, these factors also carried great weight through development.

While the benefits of a blockchain-based system are clear from the literature, the approach also presents some unique challenges. Dealing with the limitations of storage capacity are an important requirement of success, and doing so without compromising the performance of the system is a significant challenge [4]. Additionally, the incompatibility of the unavoidable GDPR legislation with blockchain adds complications to the system design [24]. This highlights how the importance of user anonymity has come to shape the direction of the project. Statistical aggregation of user data and one-way cryptography for inputs of sufficient unpredictability have been highlighted as appropriate anonymisation techniques.

The most secure and reliable way to identify users clearly comes from biometric data [81]. This is also decentralised in its nature since it does not rely on a issuing body. Therefore, it has been integrated into the identity management system, and systems to verify the data have been developed. More specifically, a neural network capable of accepting and interpreting the huge amounts of complex data retrieved from a biometric marker while remaining compatible in its storage requirements with a blockchain-based database [89].

Chapter 3

Requirements and Design

3.1 Requirements

To direct the development of the system and reasonably evaluate its performance, clear requirements were defined. The basic expectations of any digital identity management solution have been summarised in Table 3.1. These ensure that the system is compliant with all necessary regulations and that the system functions for its basic tasks.

Table 3.1: Basic Digital Identity Management System Requirements.

Requirement	Notes	Reference
GDPR Compliance	Any identity management solution must adhere to the UK data protection rights.	[2]
Protection of User Data	Through encryption and system design, user data must be protected from other parties in the strongest possible way to prevent potentially harmful consequences for users.	[12]
Authentication	A process for authentication compliant with GPG 45.	[9]
P2P Verification	Users must be able to verify each other's identity.	
Account Migration and Recovery	A user must be able to recover and move their digital identity without compromising the security of the system.	
Fraud Resistance: Stolen Genuine Identity	Identity management systems must put measures in place to prevent the theft of a user's identity. This theft may come in multiple forms.	[13]
Fraud Resistance: Shared Genuine Identity	Identity management systems must put measures in place to prevent multiple users from sharing a single identity. This is a harder problem to solve than stolen identity as the true owner of the identity is complicit in the fraud.	[14]
Fraud Resistance: Modified Genuine Identity	Identity management systems must have strong integrity and prevent the data responsible for verifying its users from being modified or intercepted by an unauthorised party.	[15]
Fraud Resistance: Fake Identity	Identity management systems must put measures in place to prevent the possibility of fraudulent identities that bypass the authentication process being generated.	[15]

The system developed in this report also hopes to fulfill goals that go above and beyond the basic requirements of an identity management solution. These goals have been selected based on background reading in the area of identity management. They are summarised in the following key performance indicators:

1. Decentralisation: The degree to which the management of the developed system is decentralised through the entire usage life-cycle. Higher decentralisation is favoured as it removes power granted to single entities, improves system self-sovereignty, and improves the privacy of users [25]. It's importance has highlighted in the literature regarding blockchain and identity management systems presented in Sections 2.1 and 2.5.

2. Data Localisation: The degree to which data is localised to the user during the authentication and verification processes. Higher data localisation is preferred as this improves the security of the system and the system self-sovereignty [12]. The value of data localisation has been noted in the 'Laws of Identity' presented in Section 1.2.

3. User Anonymity: The degree to which users can remain anonymous to other users, verifiers, and system management entities. Higher anonymity is preferred as it improves the privacy for users, so least identifying information should be used where possible [12]. User anonymity is a clear requirement in GDPR [2] and has been noted as important in other solutions studied in Section 2.5.

4. System Integrity: The degree to which verifications provided by the system can be considered reliable, secure to forgery, and up-to-date. A higher system integrity is preferred as it increases the reliability and feasibility of the system. It also plays a part in ensuring user data remains private by preventing unauthorised individuals playing any part in the network. The integrity of the system is vital for success, as highlighted by other solutions Section 2.5. It is responsible for the selection of blockchain technology, as summarised in Section 2.5.

5. Accessibility: The degree to which the system is accessible to users and verifiers with respect to any resources they must have. A more accessible solution is preferred as this is required for it to become widespread and generally feasible [12]. The importance of accessibility is stressed in the 'Laws of Identity' presented in Section 1.2.

6. Portability: The ease with which the system could be modified to work with other use cases. A more portable solution is desired as identity management has far-reaching use cases for which custom-built solutions would be wasteful [12]. Portability is once again a crucial aspect of the 'Laws of Identity' presented in Section 1.2.

7. Feasibility: The ease with which the system could be adopted in the real world. Here, factors like speed, cost, and ease of use are assessed with respect to normal users. A more feasible solution is preferred as the goal of the project is to produce a practical solution for decentralised identity management. The importance of feasibility speaks for itself. With the goal of developing a practical system, the chance of it being capable of real world use is vital.

Consideration of the degree to which the system is able to meet each of these key performance indicators will form the basis of the evaluation process for the project.

3.2 Layer One Protocol Selection

The blockchain protocol used in this solution is central to the success and feasibility of the system. It is therefore vital that this is considered and well defined within the scope of the system design. As a technology still in its relative infancy, the rate of development in blockchain is very high. Therefore, a real world implementation of an identity management system on the blockchain would need to be reactive and prepared to adjust to new technology as it arrives. Where possible, the road map for development in the space has been considered in the specification of this design.

The comparative analysis of layer 1 blockchain protocols presented in Section 2.1.1 demonstrates the relative speeds achievable by each of the available protocols. It also notes the broad support for Ethereum and its adoption within the industry. As previously stated, reported transaction speeds only apply to interactions that involve the modification of data. Within the identity management system, the only smart contract interactions that must involve the modification of data are certificate assignment, certificate transfer, and certificate deletion. It is reasonable for these events to not be instantaneous since users will not be experiencing them frequently. Additionally, these involve human or computational intervention which will cause a wait time in the order of tens of minutes regardless of blockchain delays. The only interaction that is critical to remain fast is the verification of a user's digital identity. This can be read-only process, making it fast and free regardless of protocol choice.

Based on the factors above, Ethereum is a very strong choice for the identity management system. Its continued modernisation as the technology develops shows a strong trend and its unrivalled support and tried-and-tested history gives the greatest confidence in its security and reliability. As such, it was selected as the blockchain protocol for the development of this identity management system.

3.3 Identifying Information Selection

Based on the analysis of identifying information provided in Section 2.3, biometrics clearly offer the most practical and secure solution for unique identification of users. Selection on the form of biometrics that should be used is an important trade-off between the accessibility and system intergrity key performance indicators from Section 3.1. As discussed in Section 2.4, the strongest biometric data from a uniqueness perspective are given by fingerprints and retina scans [23]. However, the tools required to gather these data are not accessible to everyone, with the current iPhone notably missing a fingerprint reader [128]. Instead, facial recognition, which only relies on a photo of a person's face, can be used as a biometric marker. It is reasonable to assume all UK residents have access to a digital camera of some kind in 2022. Additionally, a face can be easily validated by comparison to the photo ID submitted by a user, a step that already occurs in the authentication process as dictated by GPG 45 [9].

However, faces do not make the strongest biometric data. This is because they can be modified, obstructed, or can simply be similar to others. Therefore, steps must be taken to improve the performance of the methodology. Firstly, on data

collection, uniformity will be achieved using a strict set of rules, similar to those around passport photos. Secondly, the data pipeline in biometric checks is designed to minimise any variation still present in the data. Finally, the biometric data is used in combination with names and dates of birth that can be verified from the user's photo ID. This means that biometric data must only be checked in the event that two individuals potentially have the same name and date of birth. This already has a low likelihood, meaning the combined probability of two individuals potentially sharing a name, date of birth, and biometric data close enough to be detected as a match will be vanishingly small. The selection of this identifying information simplifies the sign-up process by requiring only one form of photo ID and a photo of the user, strengthening the accessibility and feasibility of the system.

3.4 Identity Certification Generation

As blockchain is an immutable technology, data stored on it cannot be deleted. A footprint will always exist regardless of the construction of smart contracts in an attempt to get around this. Mutable blockchain protocols have been proposed but there is still no widely available implementation of these due mostly to a lack of need [129]. As discussed in Section 2.2, GDPR protects a user's 'right to erasure'. This requires any service that stores personal data to delete all information related to an individual on request. Therefore, personal data cannot be stored directly on the blockchain, and instead a certification that can confirm the authenticity of user-reported data must be used. The contents and construction of this certification must be selected with utility, privacy, and practicality in mind.

Encoding Identifying Information to Detect Duplicates

The selection of name, date of birth, and facial data from Section 3.3 must be encoded in the certification for each user of the system to be uniquely identifiable. The former two data are easier to handle as they are consistent. Each time a user submits their name and date of birth it will be exactly the same and take the form dictated by their official photo IDs. As established in Section 2.2, hashing of personal data is typically not considered anonymisation under GDPR regulations, therefore an alternative solution must be found. To use the methodology of combining name, date of birth, and facial data to uniquely identify individuals and prevent multiple accounts being created with the same identity, there must be a way of identifying a potential name and date of birth clash to prevent the need for a biometric check against every user in the system. As highlighted in Section 2.2, aggregating data into bins is deemed an appropriate method of anonymisation, given the original data is deleted by the handler. This approach is used here to reduce the amount of users who may be a match with an applicant and require a biomteric check.

Facial data is more difficult to encode in the certification. While the underlying biometric data is consistent, each photo taken of a person's face will be vastly different. Facial recognition is typically performed with a neural network that takes two inputs and returns either a positive or negative match. These are able to extract the important underlying biometric data that remains consistent between images.

Therefore, some form of neural network input must be stored in the certification that can be passed to the recognition model when new data needs to be checked. However, this neural network input has two important and challenging stipulations. Firstly, the original image data used to form the processed input must not be retrievable in any way to remain compliant with GDPR legislation. Secondly, since certifications are to be stored on the blockchain, the size of the data must be minimised in order to comply with the limitations outlined in Section 2.1.2. Typically, modern camera facial data is large, coming from images of the order of megabytes. The processed image data that will be the input for the recognition model must be less than 64 bytes, or just 2 Ethereum data containers.

Collision Detection

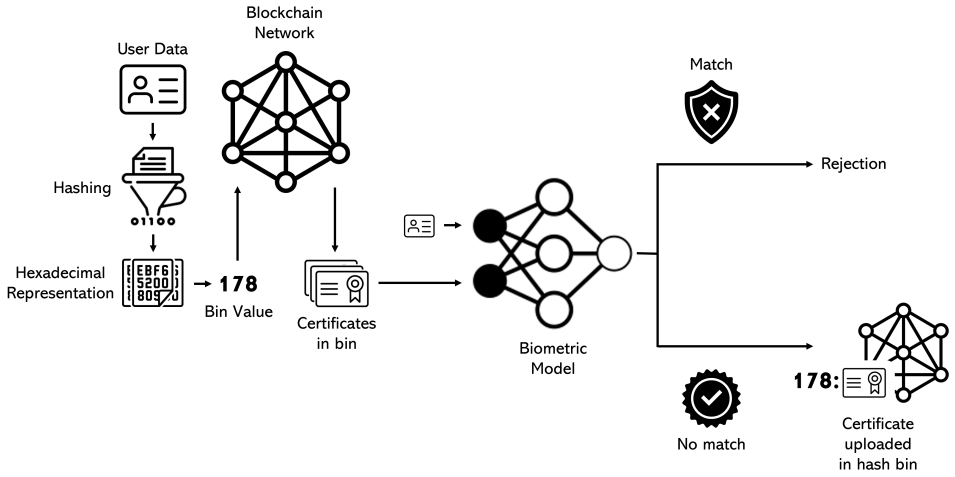

Figure 3.1: Flow chart showing how users are checked to prevent duplicate identities from being established in the system.

The proposition for this system is to hash the concatenation of a user's name and date of birth to a 32-byte hexadecimal string with SHA3-256 [130]. This has 2^{128} potential values. The hash will then be sorted into a bin based on its decimal value. The size of the bins can be selected based on the requirements of the system. A smaller bin will give greater levels of security and faster processing times, but be weaker from a privacy preservation perspective. For this solution, an average maximum bin capacity of 15 users has been selected. With total adoption of the UK's 70 million citizens [131], this requires 14.7×10^6 bins, each with 7.29×10^{31} potential values. The index of the bin a user falls into can be recorded on their blockchain certification, and on sign-up, all users in the same bin will have their biometrics checked for a match. This solution is particularly privacy protecting thanks to the addition of the hashing step. Similarity in unhashed values will be mapped to completely unrelated results, meaning two users in the same bin will not be more likely to share a name or date of birth [132]. Names are generally selected from a set of discrete values, reducing the inputs and potentially increasing the size of some bins relative to others. However, by concatenating with date of birth and hashing the output, a huge amount of variation is introduced to the system mitigating this

effect. A flow chart illustrating the process of identity collision detection on account creation is presented in 3.1.

Ensuring Data Validity

In addition to this identifying information, the certificate contains additional data to ensure that personal details reported by a user have not been changed since they were authenticated. A hash generated from a combination of their serialised image data, name, and date of birth is introduced. The use of image data in this record makes sure that this approach is GDPR compliant. The input will be thousands of bytes long and contain no words, making it immune to dictionary attacks. The value will be utterly unpredictable, particularly given that two images of the same person, taken under the same conditions, at the same time, will have different image data due to unavoidable micro-fluctuations and physical effects. That said, given a report of the correct data by a user, the verifier will be able to retrieve the same hash. Finally, an expiry date for the certification is included which is not unique or identifying, meaning it requires no further modification for GDPR compliance.

3.5 System Architecture

In Fig. 3.2, the system architecture for the identity management solution can be seen. Starting at the user, a request for a certificate in the identity management system is made by submission of personal details and a form of identification. The server forwards this to a third party for approval according to GPG 45. When this is complete, the server checks for any potential clashes on the blockchain by querying it with the bin of the hash of the user's name and date of birth. If a hit is detected, the image data is returned for a biometric check by a neural network on the server. Once authenticated, the sign-up server sends a certification for the user to the blockchain which can be queried by verifiers to confirm the user's identity.

The user can also make requests to the server to transfer their certification to a new blockchain address which is required for use cases where addresses are device-locked and the user wishes to move to a new device. They can request the server issue a deletion of the certification from the smart contract. This will not delete all record of the data, as noted in relation to the immutability of blockchain, but it does prevent the certification being used for future verifications and allows the user to set up a new account in the future without worrying about clashes with their old account.

If an account becomes compromised, the user can report it as stolen or lost to the server. With no access to their blockchain address or private key, there is no way of directly finding their account. However, the process described to prevent multiple accounts from being created from the same identity can be used to retrieve lost accounts. The user can report their details in the same manner they would on sign-up, and the blockchain can be checked for clashes. In this case, a clash will be the account lost by the user and it can be transferred to a new address for them or deleted.

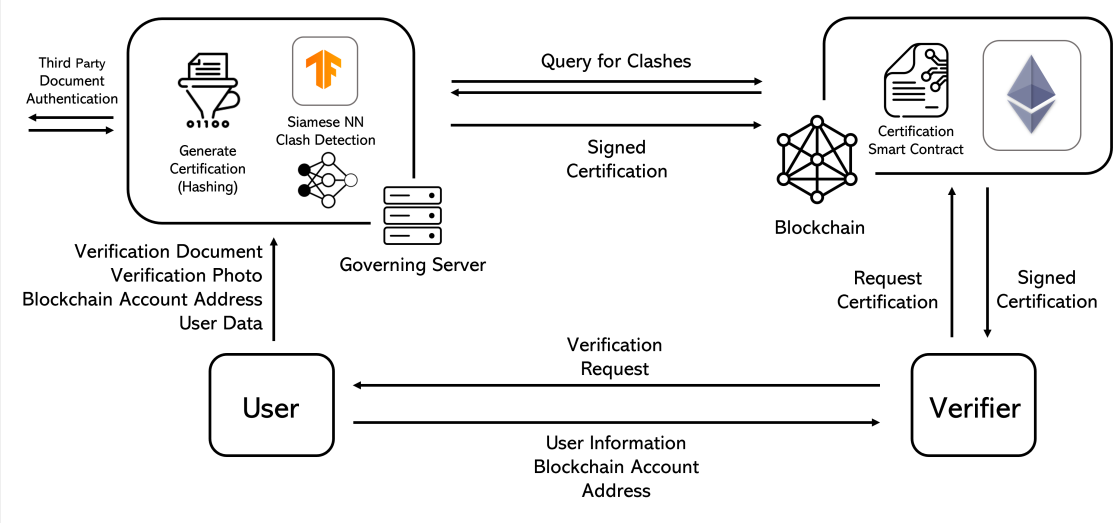

Figure 3.2: The system architecture diagram of the identity management solution. Relationships between the four main entities of the system, the user, verifier, sign-up server, and blockchain, are shown.

3.5.1 Security Protocol Design

The inter-device communications shown in Fig. 3.2 need to have proper protections in place to ensure the privacy and integrity of the system. Namely, personal data must not be exposed to the public at any point and the design must be robust to man-in-the-middle replay attacks that seek to achieve false verifications.

The deep integration of blockchain technology introduces public-key cryptography to the system. This can be used off-chain as well as on, meaning transactions at any point in the system can be signed to prove ownership of a blockchain address and to prove that tampering has not occurred. This helps prevent man-in-the-middle attacks and should be used alongside timestamping to target replays. At any point personal data are sent over the internet, it is important that proper measures are used to protect it. The system should use a protocol widely known, tested, and used for these forms of communication.

Notation

To represent the security protocols used in the system, the following notation will be used:

$$\begin{aligned} U &: \text{User} \\ V &: \text{Verifier} \\ BC &: \text{The Blockchain} \\ S &: \text{Governing server} \\ DA &: \text{Third party document authenticator} \\ PD_x &: \text{Personal Data of x} \end{aligned} \tag{3.1}$$

$$
\begin{aligned}
TS: &\quad \text{Timestamp} \\
h(x): &\quad \text{The hash of x} \\
BD_x: &\quad \text{The biometric data of x} \\
K_x: &\quad \text{The blockchain address of x} \\
K_x^{-1}: &\quad \text{The blockchain private key of x} \\
K(x): &\quad \text{Encryption of x by key K}
\end{aligned}
\tag{3.2}
$$

Authentication

In the authentication process, a user must communicate their personal details to the server. These must be passed to a third party GPG 45 authenticator, after which a certificate must be securely issued. The security protocol designed for the system's authentication process is given below:

$$
\begin{aligned}
U \rightarrow S: &\quad HTTPS\{K_U, K_U^{-1}\{PD_U, TS_1, BD_U\}\} \\
S \rightarrow DA: &\quad HTTPS\{PD_U\} \\
DA \rightarrow S: &\quad HTTPS\{PD_U\} \\
S \rightarrow BC: &\quad K_S^{-1}\{h(PD_U)\} \\
S \rightarrow BC: &\quad K_S^{-1}\{h(PD_U), BD_U\}
\end{aligned}
\tag{3.3}
$$

In the first transfer, the user sends their personal data and a timestamp signed with their private key, along with their public key. The timestamp prevents the data from being used in replay attacks and the use of signing proves that the blockchain account reported belongs to them. An HTTPS connection is used for security as this meets the requirements previously stated. The sign-up server forwards personal data to document checkers for authentication. The document authenticators return the personal data, along with an indication of acceptance. The server queries the blockchain with a hash of the personal data. Any clashes will return the biometric data stored for those accounts so checks for matches in the data can be run by the server. If no matches are found, the server submits the signed certificate to the blockchain. The blockchain will only allow the server to issue certificates and any other signatures will be rejected.

Verification

It is equally important that the verification process used by the system is secure. To achieve this, the following security protocol has been developed:

$$
\begin{aligned}
U \rightarrow V: &\quad \{K_U^{-1}\{PD_U, TS_1\}\} \\
U \rightarrow V: &\quad \{K_U\} \\
V \rightarrow BC: &\quad K_V^{-1}\{K_U\} \\
BC \rightarrow V: &\quad K_{BC}^{-1}\{h(PD), BD_U\}
\end{aligned}
\tag{3.4}
$$

In this case, the user sends their personal data and a timestamp signed with their private key, along with their public key in a separate (potentially encrypted) message. The timestamp prevents the data from being used in replay attacks and the use of signing proves that the blockchain account reported belongs to them. The verifier makes a request to the blockchain for the user's certification. Blockchain transactions are always signed. The blockchain returns the certification which contains hashes of key data that can be used to verify the individual.

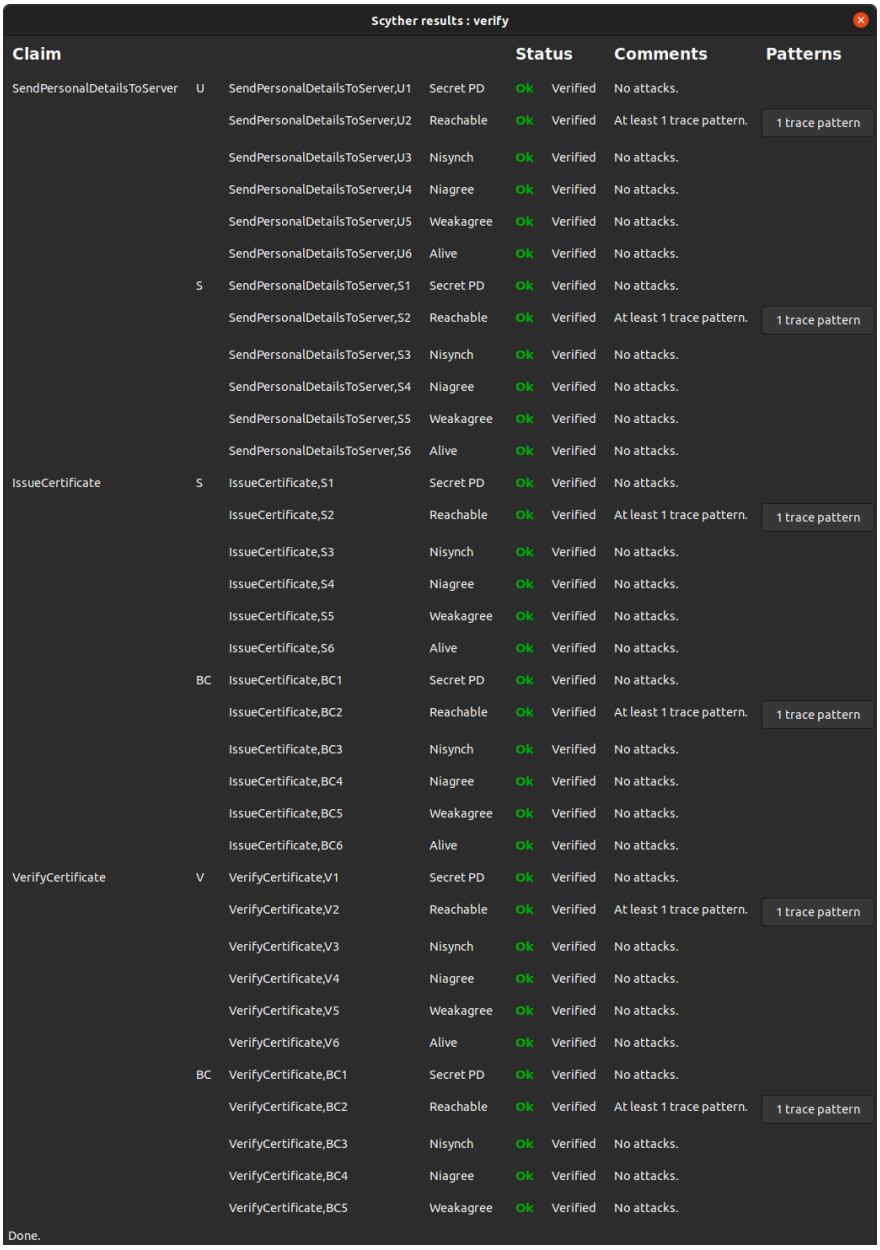

Figure 3.3: Results of the automated protocol testing carried out with Scyther. The tests confirmed that the user's identity remained secret and that the system was secure to replay attacks.

Testing

Scyther [133], an automated security protocol verification tool, was used to ensure the integrity of the protocol design. The steps of each process were entered into the graphical user interface. The protocol was then checked for the secrecy of the user's personal details and the robustness of the system to man-in-the-middle attacks. Robustness to attacks is confirmed through differing levels of synchronisation and agreement defined within the program. These evaluate the extent to which a party can be sure that an individual they are communicating with is who they claim to be and not the perpetrator of a replay attack. The results of the testing are presented in Fig. 3.3. The protocol passed on all requirements covered by the software.

3.5.2 Measures for Identity Fraud Prevention

As stated in the system requirements in Table 3.1, it is important that the identity management solution is robust to all potential forms of identity fraud highlighted in background research. Measures must be taken to combat these forms of fraud at authentication, when a user may be submitting fraudulent physical documents, and at verification, where a user may be attempting to use a false digital identity or infiltrate the identity management system. The decisions that have been made to combat the forms of fraud are given below:

- **Stolen Genuine Identity**:
  - **At Authentication**: The document checks performed in accordance with GPG 45 [9] should detect cases of stolen identity. To help the reliability of this, videos of users reading a random number should be collected instead of static images, proving the presence of the true ID document owner.
  - **At Verification**: Certifications in the digital identity management solution are linked to blockchain account addresses. To successfully verify with a stolen digital identity, the fraudster would need this account address, its private key, and all of the personal details of the true user. The potential for addresses, keys, and information to be stolen does expose additional risk to the digital solution when compared to physical ID documents, so they must be carefully handled and stored. They should not be stored together. If additional security is required, a biometric check can be performed on verification since the biometric data is stored in the certificate. This would make cases of stolen identity highly impractical, however, it would require a server call to run the neural network which would have a slightly longer delay than ordinary verifications. The digital identity management solution has the advantage of being able to easily and quickly invalidate past certifications and reissue a new one when cases of stolen identity are reported.

- **Shared Genuine Identity**:
  - **At Authentication**: The risk of multiple users trying to sign up with the same identity is notably high in a digital identity management system. The problem is compounded by the existence of multiple identity documents belonging to a single person. In the shared identity situation, an individual over the age of 18 will have their own digital identity account. They may then choose to set up an account using their details for a younger sibling or friend. They may use a different document to try to go undetected. Because the true owner of the identity is complicit, the photo submitted will match the document supplied, so the fraud will not be detected by GPG 45 [9] checks. Since the identity management system uses the combination of name, date of birth, and facial data to uniquely identify users, a clash will be detected as the combination submitted already exists in the system and the authentication will be rejected.
  - **At Verification**: There is also a high risk of users sharing their blockchain address and keys with friends or siblings to bypass the system. As with stolen identity, this can be combated by performing biometric checks in verifications of particular importance. In cases where the identity only needs to be used on a single device, such as a digital proof-of-age mobile app, blockchain account addresses and keys should never be openly accessible. By locking accounts to the devices they are used on within the software that manages them, the information cannot easily be passed on. Additionally, the software that manages these accounts should never take a user input, meaning that even if an account address and key were discovered, it cannot be used on another device.
- **Modified Genuine Identity**:
  - **At Authentication**: Documents that have been doctored for submission on sign-up should be detected by the process governed by the GPG 45 [9] checks. If additional security for a given use case is required, multiple documents can be requested.
  - **At Verification**: The digital identity managment solution is very robust to modifications. If a user is able to modify the data stored locally on their device or reports incorrect data to a verifier, the certification will not match it and the verification will fail. The certification smart contract should be constructed in such a way that only authorised parties are able make any changes to it. Additionally, transmission protocols are designed with replay and man-in-the-middle attacks in mind.
- **Fake Identity**:
  - **At Authentication**: Documents that are fake should be detected by the process governed by the GPG 45 [9] checks. Document reference numbers can be validated and false information flagged.

– **At Verification**: The certification smart contract should be carefully constructed to ensure that only authorised parties can issue certifications. Transmission protocols are designed with replay and man-in-the-middle attacks in mind.

3.6 Identity Management System in Application: Digital Proof-of-Age

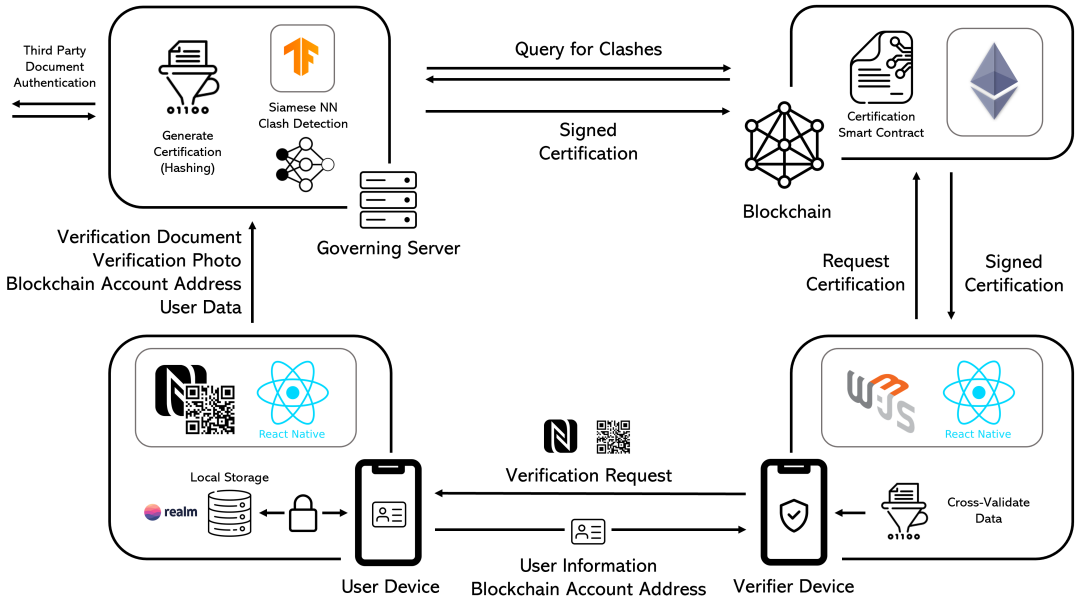

Figure 3.4: The system architecture diagram of the identity management solution and the integration of a digital proof-of-age app. This diagram shows how the system can be implemented in real world settings.

To demonstrate the function of the digital identity management system and the process by which users can be verified, a digital proof-of-age app has been developed to act as a proof-of-concept. This app is designed to be used for the purchase of age-controlled goods, so must meet a high level of integrity if it hopes to fulfill the regulations that would be put in place by the government. The app must communicate with both the sign-up server and the blockchain to fully utilise the identity management system, as shown in Fig. 3.4.

When a user wishes to be verified by a verifier, they trigger a data transfer of their identity information. This is timestamped to prevent replay and signed to prove authenticity and provide encryption. They can then communicate their blockchain address to the verifier so the data can be unsigned and the account ownership verified. Once the verifier has the user's details and blockchain account address, they can query the blockchain to retrieve the user's certification. The certification is cross-checked with the data reported by the user to ensure validity, and the result of the

verification is reported to the verifier in-app. A flow chart of the full verification process is shown in Fig 3.5.

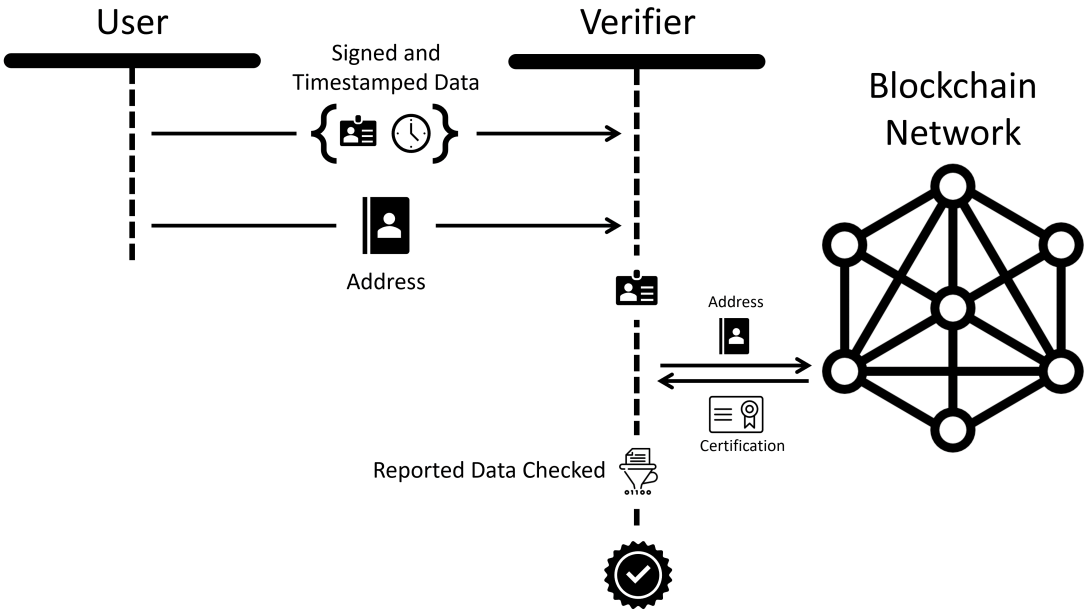

Figure 3.5: Flow chart showing how users are verified in the system.

This proof-of-concept application chose not to integrate a biometric check in the verification process. While this could be easily integrated given the availability of biometric data on the identity management system, it did not make sense in this use case. Firstly, integration of a biometric check on verification would most likely require a server call. The biometric model used in the system has relatively heavy computational requirements which would not perform well on a mobile platform. Additionally, it would not be in the best interest of security to distribute the model and reveal the weights used within it. Implementing a server call would introduce significant delays which are not acceptable from a digital proof-of-age app where the verifier is looking for rapid feedback and may be dealing with multiple users. Secondly, the biometric check would be unnecessary in this scenario. Since it is an application for use in physical settings, the verifier will have the user in front of them and a copy of their identity on their device. The face verification can be done manually in this setting to a high degree of certainty.

Therefore, the proof-of-age app only uses the biometric identification during the sign-up process to prevent duplicate accounts from being set up. Alternate use cases such as online bank applications could use the biometric identification features during account verification for an additional degree of security as supported by the identity management solution.

Chapter 4

Experimentation and Development

4.1 Biometric Identification Model Development

To prevent multiple users from creating digital identities with the same government-issued documents, biometric data is used alongside names and dates of birth to generate certifications that can be checked for collisions.

The following section details the design and development of a neural network for one-shot facial recognition that is compatible with the limitations of blockchain technology.

4.1.1 Model Performance Goals

The performance goals of the neural network depend greatly on the methodology for collision detection outlined in Section 3.4. The non-deterministic nature of neural networks and variation in the human population mean that complete confidence cannot be achieved. Additionally, the vast majority of biometric checks will be non-matches, since occasions where users are trying to commit fraud by establishing multiple accounts with the same identity will be comparatively rare. Therefore, only cases where very high degrees of certainty that a collision has occurred should be flagged.

In the case where biometric checks are also being carried out during the verification process of the identity management system, the goals would be slightly different. In this case, most inputs would be expected to be matches. Therefore, the cost of misclassification in the non-matching direction would need to be increased. This can be accomplished by using a different classification threshold that favours the matching class.

By referring to a non-match as a positive identification, and matches as a negative, the desired outcome can be achieved by prioritising a high recall. However, precision is of course still vital, as the model must be able to correctly identify matches when they occur. The best metric to capture this trade-off is F-beta score [134]. This is given

$$F_\beta = \frac{(1+\beta^2)(P \cdot R)}{(\beta^2 \cdot P + R)}$$

where P is precision, R is recall, and β is a variable selected on how many times more important recall is considered compared to precision.

Defining β is somewhat arbitrary, and over-weighting recall can cause issues in performance evaluation even in cases where it is clearly preferred. Initially, a value of 2 will be used for β to assess the performance of models in a somewhat balanced way. Further analysis can be carried out to ensure that the model meets the task's specific requirements.

4.1.2 Privacy and GDPR Compliance in the Biometric Data

The system requires some biometric data to be stored on blockchain Just as with other forms of data, this can be subject to GDPR restrictions. It is therefore important that it is stored in a way in which it is irretrievable and non-identifying to ensure compliance.

The outputs of embedding models have no meaning to humans. However, all data necessary for the neural network to complete its desired task is present. The embedding model output can be stored as long as it is irreversible.

Most neural networks are irreversible, not least because many inputs map to few outputs by necessity of the tasks they have been designed to carry out [135]. Embedding models are not directly reversible. However, they must contain some amount of information about the original input to complete their desired task. It is possible to train another neural network to reverse the embedding process based on the inputs and outputs of an embedding model [136]. This is only possible in cases where the embedded output contains a large portion of the original information present in the input.

The size limitation placed on the space for embedding data in the certification has a benefit. Embeddings are only truly irreversible when information present in the input data is lost due to the size of the output. It is simply not possible to recover a 70kB image from 64 bytes of data, else this would prove to be the most powerful compression available. Furthermore, at its best a neural network provides only assessments of the likelihood of a match, not absolute classifications. This gives confidence in the ability of the embedding model to preserve the privacy of users whose data are stored on the blockchain and ensures GDPR compliance.

4.1.3 Training Data

To train and test the biometric identification model, a dataset of appropriate images was required. These images needed to be similar in composition to the images that would be received in a deployed version of the identity management system. The VGGFace2 dataset was selected for this task [137]. This is a dataset consisting of 3.31 million images of 9,131 individuals scraped from google images and consisting primarily of public figures. The dataset covers the same subjects in a broad set of poses, expressions, and ages. It makes no claims about the balance of ethnicities within the dataset and warns of the potential for unintentional biases to arise. However, it does include examples of many different ethnicities.

24,642 images were selected from the dataset for use in the model. These were particularly well suited for the task in their pose and expression. Examples of these are shown in Fig. 4.1. From these images, 46,574 samples of pairs of images were generated, with 80% not matching and 20% matching. The imbalance was introduced to increase the cost of misclassification of matches in the interest of meeting the goals expressed in Section 4.1.1. In testing, it was found that more extreme imbalances resulted in undesirable model behaviour where loss was minimised by classifying all examples as not matching.

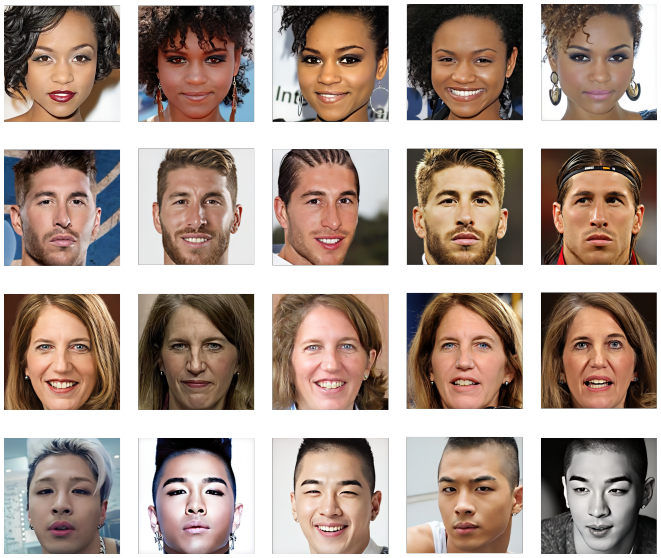

Figure 4.1: Examples taken from the VGGFace2 dataset [137]. Images of the same individual cover a variety of expressions, accessories, and ages.

4.1.4 Minimising Inconsistency in Face Data

Unlike other biometric markers, the human face undergoes significant changes. Some of these changes are natural, such as facial hair growth and aging, while others are synthetic, such as makeup and piercings. While the ease of collection makes facial recognition the only viable option for a system such as this, in every other sense it falls short of fingerprints, retina scans, or 3D facial scans. It is therefore vital that the facial recognition model is trained in a way such that features unaffected by these changes are used for detection of a match.

The change in a face due to aging is the simplest problem to overcome. In this instance, the model can be trained on a dataset that includes multiple images of the same people taken over long periods of time. As discussed in Section 4.1.3, this data is available. Further protection against aging can be provided by programming expiry dates into the digital identity certifications. Much like physical forms of identification, one would have to reapply at the end of a fixed length of time. This is good practice regardless, as it ensures that the image and information stored on the user's end are up to date.

Synthetic disruptions can be dealt with using non-technical solutions. Subjects can be asked to take photos without makeup, glasses, or piercings. This is likely required to successfully authenticate them as they will need to be compared to the image on their government-issued document. By requiring authentication before attempting to sign a person up, their image is confirmed to be of the correct specifi-

cation for the face recognition model.

Facial hair is more challenging to deal with. The model can once again be trained on data that has the same subjects with different facial hair. However, there is so much variety in the nature of facial hair that it would be close to infeasible to rely on this as a solution. Instead, the model can be helped by forcing it to pick features that will not be disrupted by facial hair. This can be done by cropping images to the area around subjects' eyes and nose. For this, an object detection model to get the correct image dimensions is used. In the interest of performance, a model from the TensorFlow 2 Object Detection Model Zoo [138] was selected for this task and transfer learning was used to train it to pick out the desired region of the human face.

Image Cropping Model Development

Figure 4.2: Examples of data used to train the object detection model for image cropping.

Of the available models in the TensorFlow 2 Object Detection Model Zoo [138], the EfficientDet D0 512x512 [139] was deemed most suitable due to its high performance on the COCO example dataset [140] and computational efficiency, reducing the strain on the servers that will be required to run these classifications. 40 labeled images from the VGGFace2 [137] dataset generated using LabelImg [141] were used in training. Examples of these can be seen in Fig. 4.2. When 100 further unlabelled images from the VGGFace2 dataset were passed through the model, all were correctly identified with a very small margin of area. The model returns the coordinates of the top right and bottom left of a rectangle encompassing the area. These coordinates are then used to crop the image and save it in its new form.

The facial recognition pipeline within the sign-up server with measures in place to deal with inconsistencies in images can be seen in Fig. 4.3. This shows the full process completed to authenticate an individual and issue them a certificate on the blockchain.

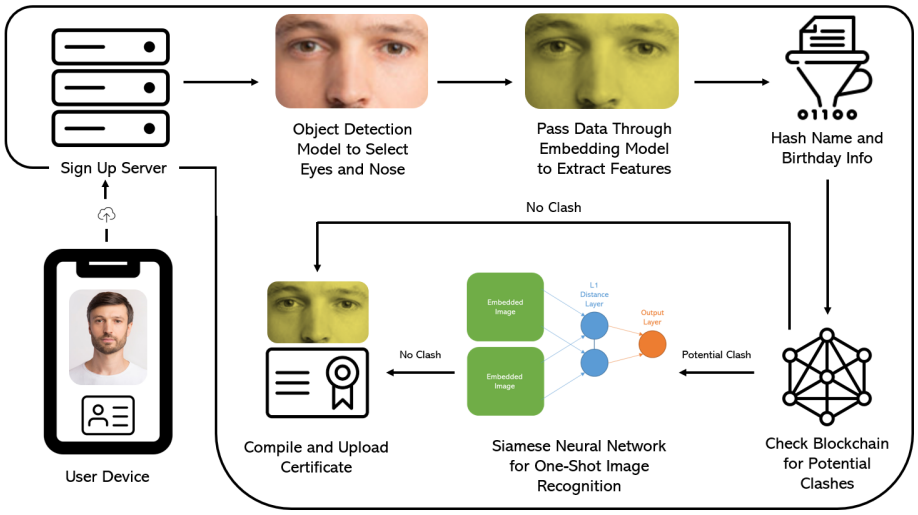

Figure 4.3: A closer look at the data processing within the sign-up server. The process a user's data goes through depending on whether or not a potential clash is detected is shown.

4.1.5 Facial Recognition Model Architecture

The facial recognition model used in this system was inspired by Gregory Koch et al's paper, 'Siamese Neural Networks for One-shot Image Recognition' [3]. This details a model architecture that takes two inputs and returns either a one or a zero to signify a match or not. Once trained, the model is capable of detecting matches between two inputs it has never seen before. This makes it perfect for the task at hand where the network will be required to perform one-off biometric checks.

The architecture of a Siamese Neural Network is shown in Fig. 2.1. It consists of an embedding model and a classification model. Both images are passed through the embedding model. Then, the outputs of this are passed together into the classification model. The network presented by Gregory Koch et al was not designed specifically for facial recognition, and was instead applied to symbols. Additionally, it was not designed with such a specific use case with respect to storage in mind. In the following sections, modifications made to make the model as suitable as possible for the identity management solution are presented.

Throughout development, optimisation, and training, a 7:1:2 training-validation-test data split of the 46,574 samples generated for the project was used. The model loss was evaluated with binary cross-entropy to minimise the misclassifications across both classes [142]. Adam was selected to optimise the model [143]. This choice was favoured for its ability to speed the stochastic gradient descent process by adapting over the learning process and modifying the learning rate. It is widely considered the best of the adaptive optimisers for most use cases [144].

4.1.6 Image Embedding Model

The image embedding model extracts features from input images. In training, it learns which features are the most distinctive, removing the need for human input to make judgements on how the classification should be done. The construction of the embedding model originally proposed by Gregory Koch et al [3] is shown in Fig. 2.1. This model was the result of extensive hyperparameter optimisation which covered all reasonable values for convolutional filters (16-256) and kernel sizes (3-20). The architecture in Fig. 2.1 outputs a tensor with length 4,096. When serialised, this takes up around 40kB of storage. As highlighted in Section 2.1.2, blockchain technology cannot be used to store large quantities of data. The size limit for the embedding data has been set to 64 bytes. This is a considerable decrease from the output of the baseline model.

The functionality of the embedding model has changed notably. Originally, its task was to extract the features selected in training from the input image. It must still achieve this goal, but now it has the added concern of doing so in the most efficient way possible as the number of features it can pass on to the classifier are limited.

Image Embedding Model Complexity

With such a small output, it was deemed likely that the model would suffer from underfitting. This occurs when a model is not complex enough to accurately capture the high-order relationships present in the data [145]. Therefore, measures typically used to combat underfitting in neural networks were introduced to the model design to try and combat this. The most important factor in reducing underfitting is model complexity [146]. Introducing more layers and neurons to the model makes it able to capture more challenging features effectively. This will, however, increase the time it takes to train the model and run it at a later date.

Fig. 4.4 presents the performance of three models on a validation dataset through training. Models with 6, 8, and 10 layers present in the embedding model are included. These were selected from a pool of 21 models that were trained with different sets of hyperparameters so that the best-performing model of each number of layers could be compared. Converse to initial expectations, increasing the number of layers present in the model did not substantially improve the performance. It is theorised that this is because learning complex features in this setting is unproductive, as the information cannot be effectively encoded in the limited embedding output. Therefore, the information is mostly lost and not passed on for classification. A shallow model, capable of only learning low-order relationships, is better suited as it can represent the features of the input in the limited output. From Fig. 4.4, the 6-layer model offers equally strong performance at a lower computational cost in training and use. Therefore, 6 layers (3 convolutional and 3 max-pooling) are used in the biometric model.

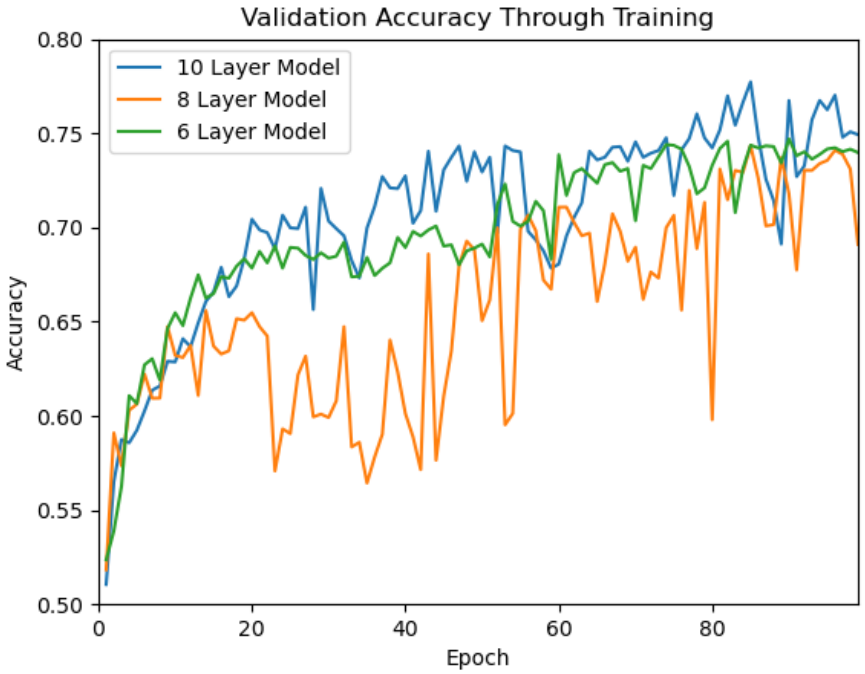

Figure 4.4: A closer look at the data processing within the sign-up server. The process a user's data goes through depending on whether or not a potential clash is detected is shown.

Image Embedding Model Hyperparameter Tuning

To go from the input image to the embedded output, a great deal of downsampling must be done. In this architecture, there are two ways in which downsampling can occur. The first is in the max-pooling layers. The entire purpose of these layers is to reduce the amount of data by selecting only the largest value in each patch of the feature map, where the size of the patch is defined by the pool size. By increasing the pool size, the max-pooling will have a greater downsampling effect as a greater number of cells will be reduced to a single cell. The second dynamic by which the input is downsampled is in the convolutional layers. Here, the kernel size can have an impact on the output. The edges of input will be lost as a result of the kernel being unable to leave the feature map. Hence, a larger kernel will result in a greater decrease in the amount of data present. This effect is particularly notable for smaller inputs where the kernel will make up a greater fraction of the feature map size. Given these two downsampling effects, kernel size and pool size are likely to have the greatest impact on the performance of the model, relative to the baseline parameters provided by Gregory Koch et al [3].

Hyperparameter tuning was carried out using a random search methodology. This was selected with computational limitations in mind to avoid an exhaustive search of the problem space. Filters in the convolutional layers were set to values between 64 and 512 in factors of 2 and kernel sizes were allowed to vary between 3 and 8 in intervals of 1, except from the first layer, which was set to 10 in accor-

Table 4.1: Best performing model found in hyperparameter tuning

	Parameter	Value		Parameter	Value		Parameter	Value
Layer Block 1	Filters	64	Layer Block 2	Filters	128	Layer Block 3	Filters	512
	Kernel Size	10		Kernel Size	8		Kernel Size	3
	Channels	96		Channels	96		Channels	-
	Pool Size	3		Pool Size	3		Pool Size	-

dance with the design by Gregory Koch et al [3]. Max-pooling layers were given 96 channels and varied pool size in unison between 2 and 3. Learning rate was set to values between $1x10^{-3}$ and $1x10^{-5}$ in factors of ten and batch size was set to 16, 32, or 64. These parameters led 55,296 total possible configurations for the model. 10% of these were randomly selected and tested in the search. F_2 score was used to select the best performing configuration after 40 epochs of training, which was found to be the earliest point of convergence in preliminary testing.

The best performing configuration for the model is presented in Table 4.1. It had a learning rate of $1x10^{-4}$ and batch size of 64. The model achieved an F_2 score of 0.93. As expected, the top performing models all had convolutional kernel sizes and max-pooling pool sizes that differed from the model suggested by Gregory Koch et al [3]. These modifications produce a smaller output prior to the dense layer that produces the final output for the embedding layer.

4.1.7 Classification Model

Classification of the embeddings is performed with a single L1 distance layer and a fully connected dense layer to reduce the output to a single continuous number between 0 and 1. This is the same design used by Gregory Koch et al [3]. To help tune the performance of the model to the goals defined in Section 4.1.1, the threshold at which matches and non-matches are decided can be altered.

4.1.8 Evaluation and Discussion

Based on the architecture that was selected from experimentation and the parameters found in the hyperparameter search, a final facial recognition neural network could be trained. The model was given a maximum of 600 epochs with early stopping based on performance in the validation set to prevent overfitting. The decision threshold that will be used with this final model can be selected so that it better meets the requirements of the system.

Fig. 4.5 presents a plot of accuracy and recall with respect to non-matches against threshold. It is clear that there is a trade-off between the two. The goal of this model is to achieve a very high recall with respect to non-matches, but not at the cost of misclassifying matches. Therefore, the recall and accuracy of the model need to be carefully balanced. From Fig. 4.5, a threshold of 0.1 was selected as it

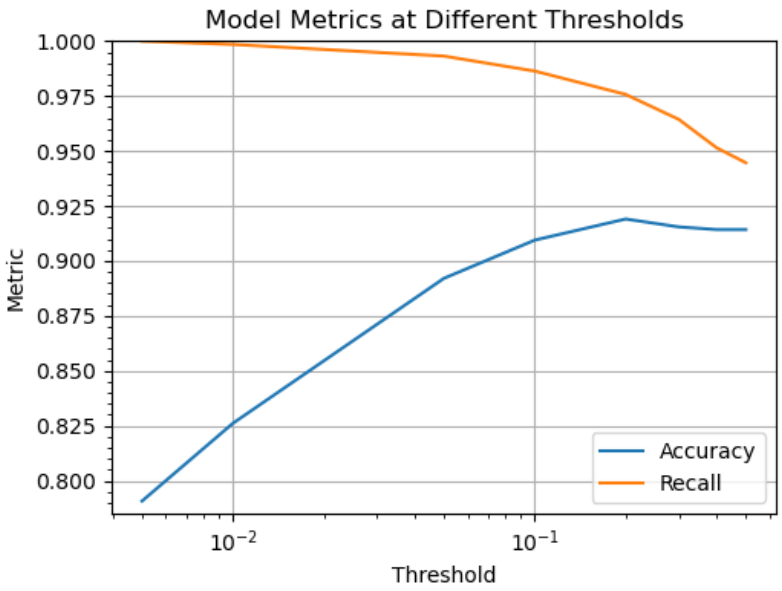

Figure 4.5: A closer look at the data processing within the sign-up server. The process a user's data goes through depending on whether or not a potential clash is detected is shown.

was deemed to suitably balance the two metrics, as can be seen in the normalised confusion matrix presented in Fig. 4.6. Effectively, this facial recognition model only declares a match when it has a very high level of certainty that the faces are the same.

The careful development of the biometric identification model has led to a performance that perfectly aligns with the needs of the system in a network that is compatible for use in a blockchain ecosystem. The final recall of the non-match class is 0.986, the precision is 0.907, and the F_2 score is 0.969. This means that non-matches are very rarely misclassified as matches, and matches are quite rarely classified as non-matches.

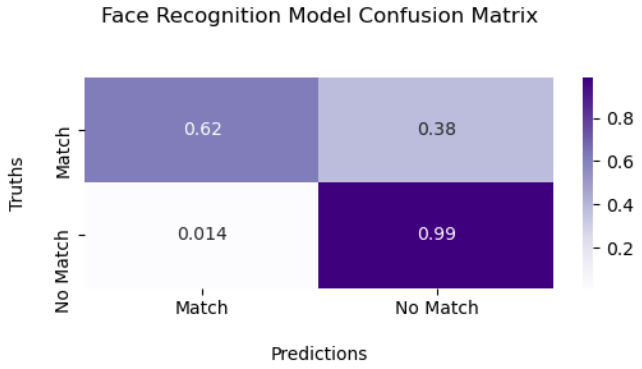

Figure 4.6: Row normalised confusion matrix of the final facial recognition model with a threshold of 0.1.

4.2 Blockchain Topology Development

As outlined in Section 2.1, blockchains come in many different flavours. While sharing the same utility at their core, the way in which decentralised consensus is reached varies based on specific use cases and differing schools of thought. Alongside the selection of a layer one protocol, the construction of the blockchain network for this system must be defined.

Based on the analysis in Section 2.1.3, node provider networks were eliminated from the selection. These removed power over the network from the governing body's and user's control. They also failed to provide any significant benefits over a custom permissioned network besides the organisational overheads required. This was not deemed to be a substantial enough reason to sacrifice control and potentially jeopardize the system.

To decide whether a custom, permissioned network or the Ethereum mainnet should be used, the identity system was deployed to both for testing of the performance, costs, and potential risks.

4.2.1 Custom Permissioned Network

The option for a custom blockchain network to host the identity management system was attractive for its ease of security management and potential for unmatched performance. Two possibilities for the topologies of a custom network were identified and one brought forward to testing to compare it with alternative options.

Potential Node Organisation Structures

The greatest challenge with proposing a custom network was in the organisation of nodes. At launch, the system would be vulnerable to attacks if a proof-of-work consensus algorithm were used since it would not have the computing power that the Ethereum mainnet has to keep it safe. This made a proof-of-authority system were nodes were selected with care the most likely option to maintain the safety of the system.

An option for every user in the system to act as a node and get a single share of the voting power in the consensus was considered. This suggestion was based on the premise that the measures put in place to prevent multiple users from sharing a single identity could be used to limit individuals to a single account. This would prevent a single person from gaining enough votes to sign blocks at their leisure. This could lead to a highly decentralised solution, with increases in the number of users in the system also making it more reliable and secure. However, this solution would potentially introduce a single point of failure for the entire network. If the system to prevent multiple accounts from being established from the same identity fails then not only are the fraudulent accounts made through these means in the system, but potentially all certifications issued could have been modified or created falsely. This solution has the additional requirement for every single user to run a device as a node. This would entail leaving it running which would incur electricity

costs and giving up storage on the device. Certifications alone could grow to take up more than 8GB in a widely adopted system.

An alternative permissioned network would take on selected parties to act as nodes. Given that not every user can act as a node, those who may must be selected with great care. A method to do this was not fully developed, but background reading from projects such a Sovrin [19] show how this can be done with trust scores derived from user-to-user interactions. There would still need to be a system of compensation for the voluntary nodes. This leaves a lot of protocols to establish in order for this approach to become feasible. This network was brought forward to testing.

Custom Permissioned Network Testing

The permissioned network topology was tested by creating a private network with Geth [147]. The network used the Clique [32] consensus method which is the Ethereum proof-of-authority protocol. This gave the system outstanding performance as no work had to be done to sign blocks and the network gas price could be set to 0. With the smart contract deployed to this network, all transactions could be completed with no cost instantaneously. The performance was impressive, but this project went no further in ironing out the many details required to get a custom permissioned network usable on a large scale with suitable security and reliability.

4.2.2 Ethereum Mainnet

The vast majority of blockchain projects do not run on their own network. They instead use a network like Ethereum and the massive degree of decentralisation and computing power it already has available. In most cases, this is far better from a security perspective than anything achievable on a smaller scale if one wishes to make the system accessible to members of the public who want to use it. For these reasons, a trial was also run to see if an Ethereum mainnet deployment could work with the system.

Initially, concerns over the privacy of data on a public, permissionless network led to this option being rejected. However, the requirement of anonymisation for the satisfaction of GDPR meant that it became a viable option given careful smart contract design.

Ethereum Mainnet Testing

To model the performance of the Ethereum mainnet, Sepolia [148], an Ethereum testnet was used. This demonstrated that through the definition of access rights, the smart contract could be made secure to efforts that may seek to modify its contents. Transactions ran through the permissionless blockchain were found to range from £1 to £4 in cost and could take up to 15 minutes to complete, meaning that measures to minimise these must be implemented to ensure that the system remains accessible and feasible.

It is possible to put such measures in place and keep the number of transactions down. It is not possible, however, to match the decentralisation, reliability, and secu-

rity of Ethereum with a custom solution. Additionally, the privacy of a permissioned network is inconsequential when the anonymity of users is considered.

4.2.3 Discussion

Based on the trials for blockchain network topology, the permissioned blockchain offers few benefits for the degree of additional work, centralisation, and insecurity it brings. Both options come with their respective costs, but Ethereum's method for resolving these are well developed. The only comparative shortfall of the mainnet is that transactions take far longer to complete. The identity management system was developed for deployment on the mainnet. The delays in transactions were considered in the development and efforts to minimise them were implemented to ensure they did not affect the user experience. Additionally, using the Ethereum mainnet gives the system the greatest possible integration with the existing Web 3 world. This will allow for the best opportunity for integration with existing and upcoming projects and keeps the system as extensible as possible to other use cases.

4.3 Verification Scheme Experimentation

In the development of the identity management system, multiple verification schemes were tested. These can be grouped into two major categories. Those that carried out verifications through a blockchain transaction, and those that did not require a blockchain transaction. Blockchain transactions take place on all nodes, and are recorded on the distributed ledger. Events that do not involve a state-change of the blockchain do not require a transaction. They are read-only so can be completed by communication with a single node to retrieve the data stored on the blockchain.

4.3.1 Blockchain Transaction-Based Verification

An identity management system that verified users with a blockchain transaction would see the data stored in a user's certification transferred to the verifier's blockchain address within the network. A test implementation of this solution was deployed to Sepolia [148], an Ethereum testnet that closely models the performance of the mainnet. Initially, this design was investigated as it would allow user's to control who had access to their certification. Additionally, they would be able to view all transactions that had taken place to access their data on the past blocks of the blockchain, giving them agency over their position in the system. Verifications took anywhere from 2 to 15 minutes to go through and incurred gas costs equivalent to £1-£4 dependent on the exact design of certifications and the smart contract running the system. These qualities are unreasonable for an identity management system. While a system like this could work for implementations that will already have long wait times and occur rarely, such as banking applications, the costs are likely too high to see any kind of widespread adoption and the waiting times make it completely unsuitable for most modern use cases. No other current layer 1 protocol available in Table 2.1 is considerably better suited to this approach of identity verification.

4.3.2 Blockchain Transaction-Free Verification

An alternative system that allows verifiers to view user's certifications in a read-only operation was formulated to try and combat the issues outlined above. Unlike the original methodology, this did not create a record of certification access. However, in this instance the certificate is completely anonymous. Therefore, no valuable user data is revealed when accessing it. The user remains in complete control with who they share their identity with as it is only with a report of their personal data that the certification can be used. For these reasons, the lack of record in data access was deemed acceptable. With this design, the verifications can be completed almost instantaneously with no cost.

4.3.3 Discussion

The verification scheme trials led to the decision to use a transaction-free model for verifying users of the identity management system. This was deemed to provide appropriate levels of user control while remaining free, fast, and easy to use. The decision aligns well with the selection of the Ethereum mainnet as the blockchain topology of choice for the system.

Chapter 5

Implementation

Implementation of the digital identity system encompassed a broad set of components. As a blockchain-based solution, there were on-chain aspects to consider and build. However, for the full system to operate, it also required a server designed to govern the system and a user device to interact with it. Through the following chapter, the implementation of these components is summarised.

5.1 Identity Management System

The identity management system is the core product of this project. It is comprised of the blockchain-based storage of user certifications and the server required to manage the system.

5.1.1 On Chain

With Ethereum chosen as the protocol for use in this project, the development of the on-chain aspects of the solution was done in Solidity, Ethereum's proprietary programming language [43]. The data stored on the blockchain is tracked and managed by a smart contract. This is held at an Ethereum address and can be interacted with using transactions encoding specific information.

Smart Contracts

The most important aspect of the development of the smart contract used to store certifications of identity is the access rights. While any user is entitled to request the certification of another in order to perform a verification, only authorised parties should be entitled to issue, move, or delete certifications. Function modifiers can be written in Solidity to design custom access rights for the methods of a smart contract. This was used to restrict alteration of certification details to the issuer. Future iterations of the identity management system could easily modify this to include additional authorised parties based on the use case, or incorporate a mutable list of authorised parties so that entities can be added or removed at will.

The immutability of blockchain makes the smart contract uneditable once deployed. It is therefore vital that it behaves as expected and does not contain any bugs or vulnerabilities. This is a big ask of any program, and to help make this possible, the amount of logic in the smart contract has been minimised as much as possible. As its primary function is to store certifications, much of its behaviour can be encapsulated with a simple mapping object. This is a hash map that is used here to map blockchain addresses to their associated certifications. With the rest of the contract revolving around modifications of this mapping, it is achievable to keep the overall amount of logic in the contract down.

Limiting the size of the smart contract has the additional benefit of minimising transaction costs for deployment and method calls, since more complex and space hungry computations use more gas.

Deployment

For the purpose of this project, the identity management system was deployed on Sepolia, an Ethereum testnet that was forked from the mainnet [148]. This was selected as it closely models the environment and performance of the Ethereum mainnet, yet it has a valueless token as the chain's coin, making smart contract deployment and data-modifying transactions free. This is ideal for the development and debugging process and functions perfectly as a proof-of-concept.

The deployment of the smart contract allows the cost of the system to be assessed. While the Sepolia token is valueless, it runs the Ethereum virtual machine and will incur the same gas cost as the Ethereum mainnet. The gas costs of the smart contract operations and an estimate of the cost in GBP, given a gas price of 10 gwei [149] and ETH price of £1,630.13 [150], are presented in Table 5.1. Due to the nature of cryptocurrency, these are of course subject to fluctuation. However, these costs are very low considering that they are the only ones for the system. Once established, an account can be used with no limit for no additional costs.

Table 5.1: Cost of operations for the smart contract.

Operation	Gas Cost	Estimated Cost (GBP)
Contract Deployment	1461627	23.80
Issue a Certificate	244419	3.98
Delete a Certificate	115649	2.83
Move a Certificate	120576	2.95

5.1.2 Off Chain

The off-chain component of the identity management system is responsible for certificate issuing, certificate transfer, and certificate deletion. These tasks are left in

the control of the server as they all involve data modification, meaning that they will incur transaction fees on the Ethereum mainnet. However, it should be noted that the anonymity of all users and transparency of the blockchain still work to minimise the control of this central authority.

Based on use case, the system could be easily modified to allow users to delete and transfer accounts themselves. This would require a level of knowledge from the users as they would need to be able to transfer some funds to their blockchain account in order for the transactions to go through. However, users cannot be given the ability to issue themselves certificates as this would not comply with the requirements of GPG 45 [9].

The server was built using Node.js [151] and Express [152] to respond to HTTPS requests from users. It interacts with the blockchain using Web3.js [153]. This hugely simplifies the process of sending transactions to the blockchain and helps ensure proper security practices when doing so.

To conform to GPG 45 [9] regulation, the server would need to make a request to a third party body that uses a government-approved process to authenticate individuals' identities. This is out of the scope of this project, and was simulated using a call to a mock server that simply returns a positive verification upon its invocation.

Deployment

To make the server available to users, it is hosted on the free tier of AWS EC2 [154]. This has limited computing power but is able to run the biometric model for face recognition. A real world solution would require more computational power in its servers to support multi-threaded requests and maintain quick processing times.

If the server becomes unavailable, the identity management system will still be able to function. While no new users can join and account transfer and deletion operations are lost, its independence from the blockchain responsible for the day-to-day activity of the system will keep most functionality running. This is crucial to maintain the reliability attributes of blockchain and remove any single points of failure.

5.2 Digital Proof-of-Age Application

The digital proof-of-age-application was built in React-Native [155]. This was selected for its cross-device compatibility. Through development, no iPhone or Macintosh computer was available to test the application, and a small amount of native code was written. Therefore, the application is not expected to be fully functioning on IOS devices. However, the vast majority of the code is transferable making the process of migration relatively simple.

Communication with the blockchain was handled using the Web3.js [153] package. The application required some configuring to make the package compatible with native environments, since it was designed for web browsers, but with this alteration the application is capable of full communication with the blockchain network. Communication with the sign-up server is handled using HTTPS requests.

These are simple to implement, widely adopted, and safe. They make a solid choice for this case in keeping complexity and the opportunity for vulnerabilities to be introduced down.

In the system, user details are only stored locally. This is handled with an encrypted Realm [156] in the mobile application. This is separate to the keystore [157] responsible for handling the user's account information. The data is not modifiable by the user. However, if they are somehow able to change the information or image that is stored, the certification will no longer be valid as it will disagree with the reported data.

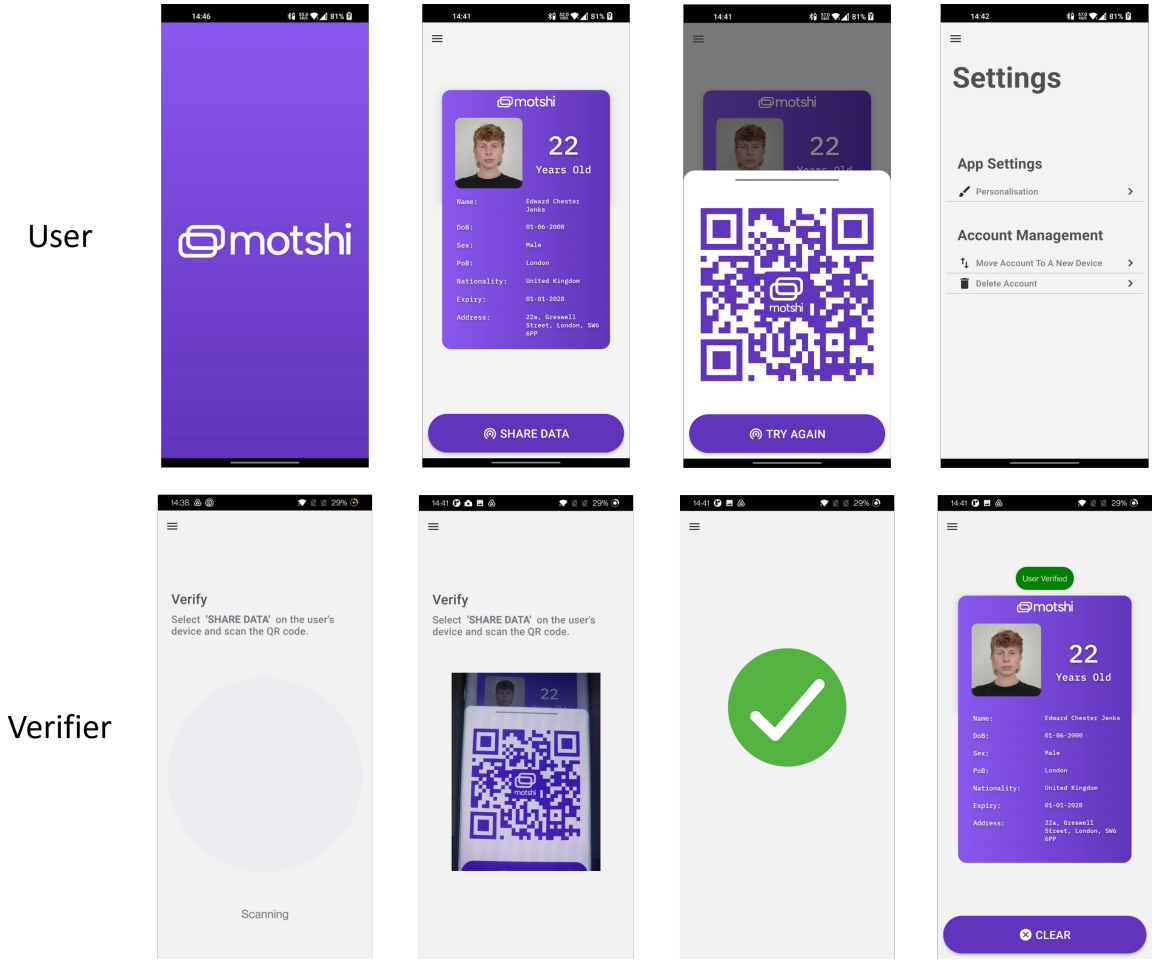

Figure 5.1: The user interface of the proof-of-age app developed alongside the identity management system.

The user interface of the application is shown in Fig. 5.1. A new user can choose to sign up, import an account from another device, or report an account stolen. The sign up process consists of data entry and upload of document and facial images. On submission, the sign-up server handles the request and issues the certificate. The device detects the issuing of this certificate on the blockchain and gives the user full access to the application's functionality. To report an account as stolen, the user follows the same data entry process. They will then be notified at a later date if their request was successful.

Once an account has been established, the user will have access to a digital representation of the ID, a page in which they can verify the identity of other users, and a settings menu that support app personalisation, account deletion, and account transfer.

Verification Process

A user can trigger a verification with the press of a button. This publishes their personal data using the Google Nearby Messages API [158]. This is a cross-platform API that uses Bluetooth, audio, and WiFi transfer to efficiently and reliably cater for local area data sharing. The user's information is timestamped and signed with their blockchain account to prove that it belongs to them and prevent replay attacks. In addition to the publishing of their data, a QR code generated from their blockchain address will be displayed on their screen.

A verifier device will be subscribed to Google Nearby Messages with the same API key, making any published data visible to them. They will then scan the QR code displayed on the user's device and use that to unsign the published message. With the personal details and blockchain account address of the user now available, the verifier can make a request for the user's certificate from the blockchain. They can then carry out the checks necessary to confirm that the user's reported information and image have not been modified, and the certification is not expired. Once verified as authentic, the user's identity can be displayed on the screen of the verifier for inspection.

Account Transfer Process

With the 1-3 year lifespan of modern smartphones, it is important that users are able to migrate accounts when they change device. Functionality to support this has been built into the proof-of-age application.

On selection of the 'Import Account' option on the sign-up selection screen, the new device will display a QR code corresponding to the blockchain address of that device. The old device will scan this address and simultaneously be publishing the personal data of the user using Google Nearby Messages. The new device will detect this data and save it. The old device can then request the sign-up server to trigger an account transfer from its address to the new address in a signed message. The process has been carefully designed such that if any single part of it fails, the entire transfer will not go through. This prevents instances where the blockchain transfer is successful but the local data transmission failed, leaving the new device and old device both locked out of the digital identity.

5.3 Integration and Testing

The system design is comprised of many separate, interacting components. It is therefore important that each individual aspect is thoroughly tested with unit tests to ensure that the expected behaviour is achieved. Additionally, the system integration

must be tested to make sure that communication between the different aspects occur as expected, and error states are handled properly in cases of failed integration.

The immutability of blockchain makes rigorous testing of the smart contract particularly important. Once deployed, it cannot be modified and the only way to make changes to the system would be to deploy a new version. This would result in the loss of all existing data in the system, requiring all users to resubmit their data. This is unacceptable in a real world implementation and must be avoided at all costs. Tests can be written for Solidity in a second smart contract that simulates real world interactions with the primary contract. The process can be simplified with Truffle [159], a DAPP development tool. This makes running tests, compiling smart contracts, and deploying to the blockchain simple and easy to reproduce. Every aspect of the smart contract's behaviour and its access permissions were tested with this methodology to ensure it works as expected.

The deployment of the on-chain and off-chain aspects of the identity management system made integration testing from the proof-of-age application repository possible. Development of adapters to deal with the communications minimised the potential for errors in integration to be introduced and narrowed the scope for testing as much as possible. Simulated interactions are run to ensure expected behaviour is observed across the entire system.

To help with the automation of the testing process, Gitlab CI/CD [160] was used to run tests every time a version of the source code was pushed to the remote repository. This helped manage the system through development and continuously ensures new bugs are not introduced as updates are developed.

Chapter 6

Evaluation

Sensible evaluation of the identity management system is required to accurately determine if the proposed solution is successful in meeting the project aim. Through careful consideration of the system specification and identification of flaws in the design, such an evaluation is presented in this chapter.

6.1 Basic Requirements

The system can be assessed against the requirements detailed in Section 3.1 to analyse whether or not it has suitably fulfilled its minimum functionality. The basic requirements of an identity management system have all been fulfilled, as shown in Table 6.1. These are essential for the system to even be considered as if it falls short of this basic functionality, it cannot be adopted in the real world.

Table 6.1: Basic Digital Identity Management System Requirements.

Requirement	**Status**	**See More**
GDPR Compliance	Achieved	Sections 3.4 and 4.1.2
Protection of user Data	Achieved	Sections 3.4 and 5.2
Authentication	Achieved	Section 3.5
P2P Verification	Achieved	Section 3.5, 3.6, and 5.2
Fraud Resistance: Stolen Genuine Identity	Achieved	Section 3.5.2
Fraud Resistance: Shared Genuine Identity	Achieved	Section 3.5.2
Fraud Resistance: Modified Genuine Identity	Achieved	Section 3.5.2
Fraud Resistance: Fake Identity	Achieved	Section 3.5.2
Account Migration and Recovery	Achieved	Sections 3.5 and 5.2

Although there can be no degree of certainty over GDPR compliance without oversight from legal professionals and potentially a judgement process, the steps taken to ensure the system remains compliant have considered the most extreme interpretations of the regulations. This is definitely the strongest place to be as a blockchain-based solution given the lack of support for the technology in the existing legislation.

Aspects of the authentication process have not been innovated upon, namely the process of confirming an individual matches the identity they have reported, as governed by GPG 45 [9]. However, the development of the system to run on blockchain has required redesign of other aspects of the authentication process. The primary role of biometric data in the solution has been for the prevention of multiple accounts being set up from the same identity. These measures would not be required in a centralised system that stored user data remotely as clashes could be detected directly.

It is fair to say that the blockchain component of the system has made the basic aspects of an identity management system more secure and easier to manage. The built-in security measures and public key infrastructure make the verification processes reliable and readily available at no cost to the consumer or system manager. This is achieved without sharing any personal user data over the internet. While the system suffers from more convoluted authentication and account management operations, the benefits of the blockchain are clear here when compared to centralised options.

The nature of any digital system makes it comparatively easy to invalidate stolen identities relative to physical identity solutions. However, this system has the additional benefit that the certifications used to verify user's identities expose no personal information about the people they represent. Centralised solutions that store user data could still lead to mass identity theft if the information is compromised, even if the digital identities are invalidated.

6.2 Key Performance Indicators

In addition to the basic requirements of an identity management system, the key performance indicators proposed in Section 3.1 must also be analysed. Each of these have been discussed below with an assessment of the degree to which they have been reached in a 10 point scale.

Decentralisation

In order to score a 10 for decentralisation, the system would need to be entirely decentralised. That is, to have no aspects of the management, maintenance, or storage reliant on a single body. Conversely, a score of 0 would represent an entirely centralised system where all aforementioned roles are reliant on a single party.

The identity management system has been assigned a score of 5 out of 10 for decentralisation. The heart of the identity management system is decentralised. This is the aspect responsible for the management of day-to-day activity of the system by

storing and reporting certifications as they are required. The decision to host this on the Ethereum mainnet has maximised the decentralisation of this aspect of the system as a similar level would be unachievable on a privately-run or provider-run network.

The decentralisation of the identity management system offers the best possible availability and reliability of data access to users, transparency over the state of users in the system and who is viewing their certification, and integrity of the system with respect to the validity of users who have been authenticated. This is available at an incomparably low cost which is utterly unachievable through alternative, centralised means. A small upfront cost on account creation is the only fee without further account modification.

The system is not completely decentralised. To some extent, this is inevitable under the current legislation. The requirements for identities to be authenticated under specific regulations such as GPG 45 [9] and the need for accountability in identity systems cannot support a fully decentralised system. Future legislation may move to make decentralisation easier to adopt in such contexts. In this instance, a system more similar to that adopted by Sovrin [19] where a user-run trust system is used to assess the authenticity of a given user's identity. Alternatively, if such possibilities do not achieve support, the system presented in this report is compatible with multiple issuing bodies. While having to comply with the same practices, this could help improve the decentralisation while maintaining the integrity of the system and offer potential users multiple options for who they share their data with.

Aspects of the system could be modified to be more decentralised. The decision was made to manage all data modifying transactions through the issuer. This was to improve the accessibility of the system so that users did not require knowledge on managing blockchain accounts or acquiring funds to use the service. A model seeking to prioritise decentralisation over this could choose to allow users to delete and transfer their own accounts should they have the knowledge. Alternatively, additional entities could be given the right to make these modifications. Unlike certificate issuing, this would not require the same levels of trust, given it is not regulated, so a broader set of parties could be accepted for this role.

Data Localisation

In order to score a 10 for data localisation, the personal data of users would never leave their personal devices. This would be true through authentication, general participation, and verification within the system. A score of 0 would represent a system where no personal data is stored locally, and instead all data storage is handled by a remote server.

This system scores a 7 out of 10 for data localisation. For the most part, the system achieves a fantastic level of data localisation, particularly in relation to centralised models. Once assigned an account, the only place that the user's data is stored is on their personal device. This gives them the maximum possible control over their data and who they share it with. Additionally, this avoids the transmission of their personal data over the internet during verifications which minimises the risk of it being intercepted and stolen by a bad actor.

Unfortunately, it is not possible to localise the data through the entire process with this style of identity management system. The user data must be submitted to a server for processing and GPG 45 [9] authentication. While this data is deleted as soon as possible, the lack of transparency that is present on the blockchain aspects of the system leave an element of trust in the issuing entity to delete data as they claim. This claim could be strengthened by placing hardware limitations that would make it impossible to store data by limiting the storage present on the server and by the issuing party subjecting themselves to regular third party audits that confirm their practises. This is not a perfect solution, however, as there would be ways to bypass both of these measures.

If the system was widely adopted, it could be modified to improve data localisation. Users could report to sign-up locations in-person where their identity could be authenticated by hand. They could then generate the certificate themselves, and pass that through the in-person authenticators to the issuer server so that it can be recorded on the blockchain. This would result in full data localisation but would incur large costs in both money and time that do not exist in the automated solution. Realistically, it is not uncommon to share personal data online in the current ecosystem, so users would likely be comfortable with a one-off transmission. In a future generation that becomes more concerned with privacy, however, a physical sign-up process to the digital system may be preferred.

User Anonymity

In order to score a 10 for user anonymity, users of the service would remain utterly anonymous in accordance with GDPR regulations through all participation in the service. A score of 0 would represent a system where the user is in no way anonymous, and their details are stored in a retrievable form.

The identity management solution scores 8 out of 10 for user anonymity. With the solution for certification generation developed for the system, there is no identifying information stored long-term in the system. This gives users a fantastic degree of anonymity as they are utterly untraceable from their publicly available data. Of course, they are not anonymous to the issuer of the certificate, though this is once again an inevitability given the nature of current legislation and identity management.

An area where the anonymity of users could be improved is in the verification process. The current solution involves the user sharing all of their personal information to be modified. In many cases, this may not be required and a stronger system may seek to minimise this in line with using the 'least identifying information' concept from the 'laws of identity' [12]. If just properties of a certain attribute are required, such as proof that a user's age is above a given figure, zero-knowledge proofs could be incorporated into the solution as they are in Sovrin [19]. A simpler solution may allow users to reveal only certain attributes as opposed to their entire identity. This was considered in development, but comes with its own set of challenges. The GDPR compliance of the solution is dependent on aggregating large amounts of data to generate the certificates. To be able to verify single attributes, this cannot be done, so remaining compliant would become a challenge.

System Integrity

In order to score a 10 for system integrity, there would be no way of bypassing the authentication and verification processes to commit fraud within the system. A score of 0 would represent a system where carrying out such acts of fraud would be trivial.

The system presented in this report scores a 9 out of 10 for system integrity. The integrity of the system is very high, particularly in comparison to centralised solutions. The use of blockchain as the storage medium for certifications makes the data practically unalterable and the security of the system second to none. A centralised system would be vulnerable to an attack that modified data on a single server. Even if backups or multiple servers were used, the lack of a consensus algorithm makes it difficult to keep the data reliably in sync.

There is some room for attacks on the system, namely a compromise of the private key of the issuing body could result in a bad actor modifying the state of the system at will. In such a scenario, the only plus side is that the entire system can be nullified and a new smart contract set up. This would mean all users would need to re-register but the system would maintain its integrity.

Accessibility

In order to score a 10 for accessibility, no special requirements in terms of equipment, possessions, or service access would be needed to join the system. A score of 0 would represent a system that required unreasonable access to resources for the average user.

The identity management solution has been awarded an 8 out of 10 for accessibility. Generally, the accessibility of the system is very good. It requires no special knowledge of blockchain technology and can easily be packaged into an application, as shown with the digital proof-of-age app. Additionally, the sign up process has been carefully designed such that only a form of photo ID and a digital camera are required. These are the bare minimum requirements for any digital identity system so to only need these in this instance is a good result.

The only failure of the system from an accessibility perspective is the requirement for an internet connection to use it. In use cases such as the digital proof-of-age application it is clear how this may be problematic if a user does not have WiFi or Cellular data connection. It would not be possible to build a system with similar levels of integrity without the requirement for internet connection. A solution could be designed which issued some form of signed certification directly to users to report alongside their data, but the potential for tampering, sharing identities, and stealing identities would be far higher. On balance, the UK has very good internet access across the country making this factor a serious issue for very few. It is not worth sacrificing the integrity of the system as that must come first in identity management, particularly with the brief for this project.

Portability

In order to score a 10 for portability, the system could be applied without any need for modification to a wide set of varied applications. A score of 0 would represent a system that could only work for the use case demonstrated in this report.

The identity management system has been awarded an 9 out of 10 for portability. Throughout the system design, decisions have been made to help make the system as portable to alternative use cases as possible. The decision to host it on the Ethereum mainnet gives the best possible opportunity to integrate the system with other Web 3 projects since anyone with access to the chain can request a user's certification, given their address. The system has focused on identity attributes traditionally found on forms of ID, but this could be easily expanded to include banking details or employment status. The overall security and the reliability of the system leave it strong enough in its verification process for almost any identity management need, whether that be regulated or not.

Feasibility

In order to score a 10 for feasibility, the system could be deployed as a real world solution as it stands. A score of 0 would represent a system that cannot, and could never be deployed in the real world.

The system has been awarded a 9 out of 10 for feasibility. The feasibility of the system depends on both the success with which it can function, and the costs associated. By deploying to Sepolia [148], the speed with which verifications can be attained has been proved and the use of a blockchain for the storage of certifications improves the reliability of the system with the removal of a single point of failure that may cause outages. It should be recognised that the Ethereum network has experienced outages in the past, and the activity of projects unrelated to the system could cause some issues [47]. However, since verifications do not require a blockchain transaction to take place, these should not impact most of the system's activity, and mostly affect sign-ups, deletions, and account migrations.

The proof-of-concept demonstrated in the proof-of-age application shows how the system can be packaged for a non-technical user, indicating that widespread adoption could be possible from the perspective of user effort requirements. This is an important aspect of feasibility, and is perhaps an area where many existing blockchain projects fail to get popular support.

The cost of the system would likely be far lower than anything achievable in a centralised system. For each account, the only fee would be a small upfront cost to issue the certificate. Any subsequent retrieval of that data is free. A centralised system would have costs to run, maintain, and upgrade servers. This model allows the operation to be incredibly lightweight, scalable, and very feasible in the real world given the lack of any setup or maintenance costs.

6.2.1 Review of Key Performance Indicators

Overall, the system scored an average of 7.9 across evaluation of its key performance indicators. This is a strong result, particularly with recognition of the limitations put in place by the necessity to meet regulations and government guidance. With this in mind, it would be highly challenging to improve upon this score further, though opportunities to do so such as the use of zero-knowledge proofs in verifications have been identified.

6.3 Probability of Failure in Clash Detection

The development of the clash detection system presented in Section 3.4 allows for an assessment of the probability of failure in this aspect of the solution. This may be in the context of an individual attempting to carry out fraud, or may involve the false accusation of fraud against a legitimate user.

6.3.1 Probability of Successful Shared Identity Fraud

The probability of a set of matching details and photos not being flagged can be calculated from biometric model performance metrics. This probability would represent the chance of a bad actor that is attempting to commit an act of shared identity fraud being successful. The value will be equal to the portion of false negatives over the sum of the portions of true positives and false negatives with respect to the matching class in model evaluation. Based on the confusion matrix presented in Section 4.6, this probability will be 0.380. Given the goal of rejections only being issued when a very high likelihood of fraud is detected, this is a strong start. Additionally, the VGGFace2 [137] dataset used in training and testing contains far more variation than would be present in real world images. It is reasonable to assert that this figure forms a lower bound for performance.

6.3.2 Probability of Wrongful Account Rejection

The probability of an authentic account setup attempt being rejected can be calculated from the system parameters and biometric model performance metrics. The assumption of maximal adoption is taken, giving the highest likelihood of failure. In this limit, the expectation value of the number of individuals per hash bin is 15, as shown in Section 3.4. The probability of wrongful rejection is then the chance of the biometric model detecting a match with one of these 15 samples, given that no true match exists. For each individual biometric assessment, the probability of a wrongful match detection will be equal to one minus the recall achieved in Section 4.1.8, which is 0.014. The chance of rejection is the probability that this event occurs at least once in all of the trials. Since this will be equal to the complement of the event never occurring, it can be calculated by raising the recall to the power of 15 and subtracting that value from 1, leaving 0.191. This value may seem high, but it is an extreme scenario. Adoption by the full population of the UK is not likely given the number of people who are too young or too old to benefit from the system. It should be capable of handling the capacity. This raises the point that at these levels it can be expected that wrongful rejections will happen. At a more conservative level of adoption, there may be an average of 3 individuals per bin. This represents one-fifth of the UK using the service. In this case, the probability of wrongful rejection would be 0.029, a much more reasonable figure given users can simply reapply.

Chapter 7

Conclusions

Over the course of this report, the design, development, implementation, and evaluation of a decentralised identity management system and proof-of-concept mobile application interface have been presented. The resulting solution offers users an unparalleled level of security, privacy, and reliability. A scenario in which user data are only stored on their personal devices and certifications are stored on a decentralised blockchain has been realised.

The final proposition is an Ethereum mainnet-based smart contract that is governed by a management server. The server is responsible for running information and biometric checks that ensure the members of the system are authentic and unique. The smart contract stores anonymous certifications of the users that are able to validate reported data, but do not reveal any information themselves. Verifiers can request the certifications from the Ethereum network to carry out checks on user-reported data, but the user is left in complete control of who they share their data with. For appropriate use cases, biometric verifications can be used to give extra confidence to the validity of identity ownership.

7.1 Technical Innovations

The three primary technical innovations required for the decentralised identity management system to function have been achieved over the course of the project. The extent to which each was met is summarised below.

7.1.1 Blockchain Compatible Biometric Identification

Development of a biometric identification model that produced outputs small enough for storage on blockchain such that it could be integrated into a decentralised identity management solution was performed through careful architecture design and hyperparameter tuning. The resultant facial recognition neural network achieved a precision of 0.907, recall of 0.986, and F_2 score of 0.969.

By carefully tuning the parameters and training methodology of the network to meet the desired set of performance goals established based on the required task it must carry out, the probability of failure in the system was minimised. The chance of

a bad actor successfully abusing the anonymity of the service to establish a duplicate identity, committing shared identity fraud, was evaluated at 0.380. The chance of an authentic user being misclassified as a fraudster in the most extreme circumstance possible was 0.191.

7.1.2 GDPR Compliant Certification System

The certification system developed over the course of this project has taken the most extreme interpretations of GDPR regulations in an attempt to ensure compliance with the law. As a side effect, users receive complete anonymity in the identity management system. The design still allows for the system to prevent duplicate accounts to function, ensuring the long-term integrity and feasibility of the offering.

7.1.3 Seamless Blockchain Integration

Through careful smart contract design and experimentation with different blockchain topologies and verification schemes, a product that offers fast and free verifications has been developed. This is central to the ability for the decentralised solution to become widely adopted as without these measures, users would select centralised services offering them. The design decision to manage accounts from a governing server allows for all concerns of blockchain account management to be removed from the user, making the solution accessible to all.

7.2 Project Success

Based on the evaluation of key performance metrics defined for the development of the system, an average score of 7.9 out of 10 for attainment of the specification was achieved. While this leaves room for improvement, it represents a very strong solution to the digital identity management challenge. In recognition of the limitations faced when working with blockchain technology within the UK's rules surrounding personal data and identities, it is a particularly good final outcome. This fact, along with the attainment of the three key technical innovations for this project to function, points to successful achievement of the overall project aim.

Central to the goals of this project was the development of a system that could see real world use on a scale large enough to support the population of the UK. The final product does very well to achieve this particular goal and demonstrates a methodology for bringing blockchain-based solutions to the general public through compromise with regulation and encapsulation of more technical aspects of the design. The same processes could be applied to a range of use cases in the interest of bringing the benefits of decentralisation to the masses.

7.3 Future Work

Opportunities for improvements to this system have been identified. The potential to further the decentralisation of the solution through support for multiple managing bodies in the smart contract, support for more blockchain account autonomy for users with technical interest, and even support for systems that can run on trust policies given government support stand out as areas of particular importance given the difficulty in achieving strong decentralisation in the system. Additionally, inspiration can be taken from other projects such as Sovrin [19], where better data localisation and anonymity is achieved through the use of zero-knowledge proofs. Future work should aim to make these improvements, along with others identified from existing proposals and novel ideas, to better the overall performance of the identity management system.

The solution proposed in this report has been designed with portability in mind. However, with only one use case tested, the ability for such a system to be applied in other situations and the effects that would have its performance and costs have not been confirmed. Therefore, future studies should aim to apply the identity management system to alternative use cases across different platforms and evaluate the impacts of such actions. In particular, modification of the design to make the system more modular should be investigated. This could help prevent users storing information they do not use and keep certification sizes on the blockchain to a minimum.

The biometric model developed in this project could be improved upon. Further optimisation could be carried out on the existing architecture with more computational resources so a wider set of configurations could be explored. Additionally, other architectures should be investigated. The general structure of the neural network used in this study still represents the cutting edge of facial recognition technology, though the exact selection of layer types and ordering vary greatly. Moreover, alternate algorithms for mapping facial features to a set of data exist [98]. Many of these do not lend themselves as well to the GDPR and output size-reduction requirements of the project since they are clearly defined in the measurements they take and thus more understandable to humans and set in their size.

An ideal implementation would have seen potential for more aspects of the management process to run on the blockchain. This would give the system better transparency over the handling of user data and remove some centralised aspects of the approach. More work covering the feasibility of running neural networks on blockchain needs to be done for this to become a reality [103]. Additionally, the scalability issues faced by existing blockchain protocols need to be addressed for such high computational demands and storage requirements to be met [161].

With the inevitable digitisation of identity in the UK [1], alternative approaches for identity management must be explored and tested to ensure that the system that becomes widely adopted offers the best possible privacy, security, and reliability for its users. Ultimately, public adoption will depend more heavily on convenience and availability than concerns over privacy. Therefore, these issues must central to any future developments and evaluations.

Bibliography

[1] Home Office. Uk digital identity attributes trust framework: updated version. Available at https://www.gov.uk/government/publications/uk-digital-identity-attributes-trust-framework-updated-version, 08 2021.

[2] UK Government. Uk gdpr. Available at https://uk-gdpr.org/, 01 2021.

[3] Gregory R. Koch. Siamese neural networks for one-shot image recognition. 2015.

[4] Ethereum Foundation. Solidity types. Available at https://docs.soliditylang.org/en/v0.8.11/types.html, 08 2022.

[5] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. Available at https://www.ussc.gov/sites/default/files/pdf/training/annual-national-training-seminar/2018/Emerging_Tech_Bitcoin_Crypto.pdf, 10 2008.

[6] World Intellectual Property Organisation. Blockchain technologies and ip ecosystems: A wipo white paper. Available at https://www.wipo.int/export/sites/www/cws/en/pdf/blockchain-for-ip-ecosystem-whitepaper.pdf, 02 2022.

[7] Rayan Tanaka. The universal basic whitepaper for the "third wave" of crypto. Available at https://opensea.io/assets/ethereum/0x495f947276749ce646f68ac8c248420045cb7b5e/18910910464212071774836220916651953824815941336455609051681234007934527078 5, 07 2021.

[8] Trung Thanh Nguyen. Axie infinity. Available at https://axieinfinity.com/, 08 2022.

[9] Home Office. How to prove and verify someone's identity. Available at https://www.gov.uk/government/publications/identity-proofing-and-verification-of-an-individual, 02 2021.

[10] Team ReadID. Nfc and the gov.uk guidelines on verifying someone's identity. Available at https://www.readid.com/blog/gov-uk-guidelines-verifying-someones-identity, 08 2022.

[11] Trust ID. Trust id. Available at https://www.trustid.co.uk/, 05 2022.

[12] Kim Cameron. The laws of identity. *Microsoft Corporation*, 11 2005.

[13] National Fraud and Cyber Crime Reporting Centre. Identity fraud and identity theft. Available at https://www.actionfraud.police.uk/a-z-of-fraud/identity-fraud-and-identity-theft, 08 2022.

[14] drinkaware. Buying alcohol. Available at https://www.drinkaware.co.uk/facts/alcohol-and-the-law/buying-alcohol, 08 2022.

[15] UK Governement. How to prove and verify someone's identity. Available at https://www.gov.uk/government/publications/identity-proofing-and-verification-of-an-individual/how-to-prove-and-verify-someones-identity, 08 2022.

[16] Jennifer Thompson. Identity thefts rise to nearly 500 victims a day. *Financial Times*, 08 2017.

[17] Make Haley. Fraudscape 2022. *Cifas*, 2022.

[18] alastair walker. Could the dark web become a fake id factory? *Insurance Edge*, 06 2022.

[19] The Sovrin Foundation. Sovrin™: A protocol and token for selfsovereign identity and decentralized trust. Available at https://sovrin.org/wp-content/uploads/Sovrin-Protocol-and-Token-White-Paper.pdf, 01 2018.

[20] Christian Lundkvist. Uport: A platform for self-sovereign identity. Available at https://blockchainlab.com/pdf/uPort_whitepaper_DRAFT20161020.pdf, 10 2016.

[21] Information Commissioner's Office. Right to erasure. Available at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-erasure/, 05 2022.

[22] Biometrics Institue. Biometrics definition. Available at https://www.biometricsinstitute.org/what-is-biometrics/, 08 2022.

[23] Biometrics Institute. Types of biometrics. Available at https://www.biometricsinstitute.org/what-is-biometrics/types-of-biometrics/, 08 2022.

[24] Panel for the Future of Science and Technology. Blockchain and the general data protection regulation. *European Parliment*, 07 2019.

[25] Dejan Vujicic, Dijana Jagodic, and Sinisa Randic. Blockchain technology, bitcoin, and ethereum: A brief overview. *17th International Symposium INFOTEH-JAHORINA (INFOTEH)*, 03 2018.

[26] L. Jean Camp and Debin Liu. Proof of work (cannot, can, does currently) work. 08 2007.

[27] Savva Shanaev, Arina Shuraeva, and Mikhail Vasenin. Cryptocurrency value and 51% attacks: Evidence from event studies. *The Journal of Alternative Investments*, 3:65–77, 2020.

[28] Bahareh Lashkari and Petr Musilek. A comprehensive review of blockchain consensus mechanisms. *IEEE Access*, 9, 04 2021.

[29] Fahad Saleh. Blockchain without waste: Proof-of-stake. *The Review of Financial Studies*, 34: 1156–1190, 2021.

[30] Anatoly Yakovenko. Solana: A new architecture for a high performance blockchain v0.8.13. Available at http://gumhip.com/wp-content/uploads/2021/05/Solana-Whitepaper.pdf, 05 2022.

[31] Shashank Joshi. Feasibility of proof of authority as a consensus protocol model. 08 2021.

[32] Péter Szilágyi. Eip-225: Clique proof-of-authority consensus protocol. *Ethereum Improvement Proposals*, 03 2017.

[33] Christine Helliar, Louise Crawford, Laura Rocca, and Claudio Teodori. Permissionless and permissioned blockchain diffusion. *International Journal of Information Management*, 54, 10 2020.

[34] Go Ethereum. Private networks. Available at https://geth.ethereum.org/docs/interface/private-network, 08 2022.

[35] Disha Sinah. Top 10 cryptocurrencies with a high transaction speed in 2022. Available at https://www.analyticsinsight.net/top-10-cryptocurrencies-with-a-high-transaction-speed-in-2022/, 08 2022.

[36] Victor Ugochukwu. Binance smart chain vs ethereum: Who will be crowned the best platform for decentralised applications? Available at https://swissborg.com/blog/bsc-vs-eth, 08 2022.

[37] Algorand. How algorand is building a scalable blockchain ecosystem. Available at https://www.algorand.com/resources/blog/algorand-building-scalable-sustainable-blockchain-ecosystem, 08 2022.

[38] Crypto Vantage. How algorand is building a scalable blockchain ecosystem. Available at https://www.algorand.com/resources/blog/algorand-building-scalable-sustainable-blockchain-ecosystem, 08 2022.

[39] CoinDesk. Near protocol. Available at https://www.coindesk.com/learn/what-is-near-protocol-and-how-does-it-work/, 08 2022.

[40] Tim Beiko. Gas and fees. Available at https://ethereum.org/en/developers/docs/gas/, 07 2022.

[41] DappRadar. 2020 dapp industry report. 2022.

[42] Michael McCallum. Ethereum virtual machine (evm). Available at https://ethereum.org/en/developers/docs/evm/, 08 2022.

[43] Ethereum Foundation. Solidity. Available at https://docs.soliditylang.org/en/v0.8.16/, 08 2022.

[44] Andrew Miller, James Litton, Andrew Pachulski, Neal Gupta, and Dave Levin. Discovering bitcoin's public topology and influential nodes. *Coinscope*, 08 2022.

[45] Damiano Di Francesco Maesa, Paolo Mori, and Laura Ricci. Blockchain based access control. *IFIP International Conference on Distributed Applications and Interoperable Systems*, 05 2017.

[46] Etherscan. Ethereum blocck explorer. Available at https://etherscan.io/blocks, 08 2022.

[47] Binder Matt. Bored ape yacht club caused ethereum fees to soar to astronomical levels. Available at https://mashable.com/article/ethereum-gas-fees-skyrocket-bored-ape-yacht-club-otherside-nft-launch, 05 2022.

[48] Deen Newman. Blockchain node providers and how they work. Available at https://www.infoq.com/articles/blockchain-as-a-service-get-block/, 04 2021.

[49] Free gas networks. Available at https://besu.hyperledger.org/en/stable/HowTo/Configure/FreeGas/, 04 2021.

[50] Evgeny Konstantinov. The ethereum cloud vs. on-premises nodes conundrum. *Chainstack*, 09 2019.

[51] Joe Jaoude and Raafat Saade. Blockchain applications – usage in different domains. *IEEE Access*, 7, 2019.

[52] Michael Nofer, Peter Gomber, Oliver Hinz, and Dirk Schiereck. Blockchain. *Business Information Systems Engineering*, 59:183–187, 2017.

[53] Marcy Gordon. Dc sues zuckerberg over cambridge analytica privacy breach. Available at https://apnews.com/article/technology-political-scandals-government-and-politics-mark-zuckerberg-1dedacd5e710e9408e5fc163e7d87666, 05 2022.

[54] Scott Ikeda. Instagram breach exposes personal data of 49 million users. Available at https://www.cpomagazine.com/cyber-security/instagram-breach-exposes-personal-data-of-49-million-users/, 06 2019.

[55] GDPR. Differences between the uk-gdpr and the eu-gdpr regulation. Available at https://www.gdpreu.org/differences-between-the-uk-and-eu-gdpr-regulations/, 08 2022.

[56] EU. Recital 26. *EU GDPR*, 05 2018.

[57] EU. Article 4: 'definitions'. *EU GDPR*, 05 2018.

[58] UCL. Anonymisation and pseudonymisation. Available at https://www.ucl.ac.uk/data-protection/guidance-staff-students-and-researchers/practical-data-protection-guidance-notices/anonymisation-and, 08 2022.

[59] EU. Art. 24: Responsibility of the controller. *EU GDPR*, 05 2018.

[60] ARTICLE 29 DATA PROTECTION WORKING PARTY. Opinion 4/2007 on the concept of personal data. Available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf, 06 2007.

[61] ARTICLE 29 DATA PROTECTION WORKING PARTY. Opinion 05/2014 on anonymisation techniques. Available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf, 04 2014.

[62] Simon Schwerin. Blockchain and privacy protection in the case of the european general data protection regulation (gdpr): A delphi study. *The Journal of British Blockchain Association*, 1: 1–77, 07 2018.

[63] Stan Sater. Blockchain and the european union's general data protection regulation: A chance to harmonize international data flows. *SSRN Electronic Journal*, 01 2017.

[64] Dirk A. Zetzsche, Ross P. Buckley, and Douglas W. Arner. The distributed liability of distributed ledgers: Legal risks of blockchain. *University of Illinois Law Review, 2017-2018, Forthcoming, University of Luxembourg Law Working Paper No. 007/2017, Center for Business Corporate Law (CBC) Working Paper 002/2017, University of Hong Kong Faculty of Law Research Paper No. 2017/020, UNSW Law Research Paper No. 17-52, European Banking Institute Working Paper Series 14*, 08 2017.

[65] Luis-Daniel Ibáñez, Kieron O'Hara, and Elena Simperl. On blockchains and the general data protection regulation. *EU Blockchain Observatory Forum*, 03 2018.

[66] Florian Martin-Bariteau. Blockchain and the european union general data protection regulation: The cnil's perspective. *Blckchn.ca Working Paper Series, 2018-01*, 10 2018.

[67] Home Office. Digital identity document validation technology (idvt). Available at https://www.gov.uk/government/publications/digital-identity-document-validation-technology-idvt, 12 2021.

[68] UK Government. Passport advice and complaints. Available at https://www.gov.uk/passport-advice-line, 08 2022.

[69] UK Government. View or share your driving licence information. Available at https://www.gov.uk/view-driving-licence, 08 2022.

[70] UK Government. Identity cards. Available at https://www.gov.uk/identitycards, 08 2022.

[71] UK Government. Your national insurance number. Available at https://www.gov.uk/national-insurance/your-national-insurance-number, 08 2022.

[72] UK Government. What national insurance is for. Available at https://www.gov.uk/national-insurance/what-national-insurance-is-for, 08 2022.

[73] UK Government. Proof of identity. Available at https://www.gov.uk/government/publications/proof-of-identity-checklist/proof-of-identity-checklist, 08 2022.

[74] Charles Clarke. Identity cards bill speech. Available at https://www.theyworkforyou.com/debates/?id=2006-03-21b.181.2, 08 2022.

[75] Thales. National id cards: 2016-2022 facts and trends. Available at https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/identity/2016-national-id-card-trends, 08 2022.

[76] Financial Conduct Authority. Financial services register. Available at https://register.fca.org.uk/s/, 08 2022.

[77] NHS. What is an nhs number? Available at https://www.nhs.uk/using-the-nhs/about-the-nhs/what-is-an-nhs-number/, 08 2022.

[78] Metropolitan Police. What happens after you report a crime? Available at https://www.met.police.uk/advice/advice-and-information/acr/after-you-report-a-crime/, 08 2022.

[79] National Cyber Security Centre. Device security guidance. Available at https://www.ncsc.gov.uk/collection/device-security-guidance/policies-and-settings/using-biometrics, 08 2022.

[80] Nivetha Ramachandran. Biometric security system using the smart system concept. *International Journal of Engineering Research*, 2:3, 09 2012.

[81] DHIR, VIJAY, SINGH AMARPREET, KUMAR RAKESH, and SINGH GURPREET. Biometric recognition: A modern era for security. *International Journal of Engineering Science and Technology*, 2, 08 2010.

[82] Lucjan Hanzlik and Miroslaw Kutylowski. *ePassport and eID Technologies*, pages 81–97. 01 2021. ISBN 978-3-030-10590-7. doi: 10.1007/978-3-030-10591-4_5.

[83] Murray Scott, Seamus Hill, Thomas Acton, and Martin Hughes. *Biometric Identities and E-Government Services*. 01 2006. doi: 10.4018/9781591407997.ch009.

[84] Sangramsing N. Kayte. Dna biometric. *Journal of VLSI Signal Processing*, 5:3, 11 2015.

[85] Mark Burge and Wilhelm Burger. *Ear Biometrics*, pages 273–285. 04 2006. ISBN 978-0-387-28539-9. doi: 10.1007/0-306-47044-6_13.

[86] Abdul Matin, Firoz Mahmud, Syed Zuhori, and Barshon Sen. Human iris as a biometric for identity verification. pages 1–4, 12 2016. doi: 10.1109/ICECTE.2016.7879610.

[87] Jarina Mazumdar. Retina based biometric authentication system: A review. *International Journal of Advanced Research in Computer Science*, 9:711–718, 02 2018. doi: 10.26483/ijarcs.v9i1.5322.

[88] A. Suganya and M. Sivitha. A new biometric using sclera vein recognition for human identification. 09 2015. doi: 10.1109/ICCIC.2014.7238324.

[89] Massimo Tistarelli and Enrico Grosso. Identity management in face recognition systems. volume 5372, pages 67–81, 05 2008. ISBN 978-3-540-89990-7. doi: 10.1007/978-3-540-89991-4_8.

[90] Sotiris Malassiotis, Niki Aifanti, and Michael Strintzis. Personal authentication using 3-d finger geometry. *Information Forensics and Security, IEEE Transactions on*, 1:12 – 21, 04 2006. doi: 10.1109/TIFS.2005.863508.

[91] Faridah Yahya, Haidawati Nasir, Kushsairy Kadir, Sairul Safie, Sheroz Khan, and Teddy Gunawan. Fingerprint biometric systems. *Trends in Bioinformatics*, 9:52–58, 09 2016. doi: 10.3923/tb.2016.52.58.

[92] Jeffrey Boyd and J.J. Little. Biometric gait recognition. volume 3161, pages 19–42, 01 2003. ISBN 978-3-540-26204-6. doi: 10.1007/11493648_2.

[93] Hesham Hashim, Shatha Baker, and Ahmed Nori. Biometric identity authentication system using hand geometry measurements. *Journal of Physics Conference Series*, 1804:12144, 01 2021. doi: 10.1088/1742-6596/1804/1/012144.

[94] Ala Alariki, Sayed Alavy, Mohammad Reza Yousufi, Mohammad Aziz, and Christine Murray. A review study of heartbeat biometric authentication. *Journal of Computers*, pages 936–947, 01 2018. doi: 10.17706/jcp.13.8.936-947.

[95] Oyeleye Akinwale, Fagbola M., Ronke Babatunde, and Adebisi Baale. An exploratory study of odor biometrics modality for human recognition. 01 2013.

[96] Raul Sanchez-Reillo, Belen Fernandez-Saavedra, Judith Liu-Jimenez, and Carmen Sánchez Ávila. Vascular biometric systems and their security evaluation. pages 44 – 51, 11 2007. ISBN 978-1-4244-1129-0. doi: 10.1109/CCST.2007.4373466.

[97] Amjad K. and Sreeramana Aithal. Voice biometric systems for user identification and authentication – a literature review. *International Journal of Applied Engineering and Management Letters*, pages 198–209, 04 2022. doi: 10.47992/IJAEML.2581.7000.0131.

[98] Steve Lawrence, C Giles, Ah Tsoi, and Andrew Back. Face recognition: A convolutional neural network approach. *Neural Networks, IEEE Transactions on*, 8:98 – 113, 02 1997. doi: 10.1109/72.554195.

[99] Hamsa Abdulkareem. Fingerprint identification system using neural networks. *Nahrain University, College of Engineering Jou rnal (NUCEJ)*, 15:234, 09 2012.

[100] Qingqiao Hu, Siyang Yin, Huiyang Ni, and Yisiyuan Huang. An end to end deep neural network for iris recognition. *Procedia Computer Science*, 174:505–517, 01 2020. doi: 10.1016/j.procs.2020.06.118.

[101] Adel Saleh, Mohamed Abdel-Nasser, Md. Mostafa Kamal Sarker, Vivek Kumar Singh, Saddam Abdulwahab, Nasibeh Saffari, Miguel García, and Domenec Puig. Deep visual embedding for image classification. 02 2018. doi: 10.1109/ITCE.2018.8316596.

[102] Brenden Lake, Ruslan Salakhutdinov, and Joshua Tenenbaum. The omniglot challenge: a 3-year progress report. *Current Opinion in Behavioral Sciences*, 29:97–104, 10 2019. doi: 10.1016/j.cobeha.2019.04.007.

[103] Jae-Yun Kim and Soo-Mook Moon. Blockchain-based edge computing for deep neural network applications. pages 53–55, 10 2018. ISBN 978-1-4503-6598-7. doi: 10.1145/3285017.3285027.

[104] Nada Sallami, Ali al, Sarmad Al Aloussi, and Alousi Cis. Load balancing with neural network. *International Journal of Advanced Computer Science and Applications*, 4, 09 2013.

[105] Guojing Cong, Giacomo Domeniconi, Chih-Chieh Yang, Joshua Shapiro, Fan Zhou, and Barry Chen. Fast neural network training on a cluster of gpus for action recognition with high accuracy. *Journal of Parallel and Distributed Computing*, 134, 08 2019. doi: 10.1016/j.jpdc.2019.07.009.

[106] Daniel Coquelin, Charlotte Debus, Markus Götz, Fabrice Lehr, James Kahn, Martin Siggel, and Achim Streit. Accelerating neural network training with distributed asynchronous and selective optimization (daso). *Journal of Big Data*, 9, 02 2022. doi: 10.1186/s40537-021-00556-1.

[107] Namecoin. Decentralized secure names. Available at `https://www.namecoin.org/resources/whitepaper/`, 2010.

[108] Vitalik Buterin. Ethereum whitepaper. Available at `https://ethereum.org/en/whitepaper/`, 2014.

[109] Joel A. Bosh. I/o digital. Available at `https://www.crunchbase.com/organization/i-o-digital`, 08 2022.

[110] BITNATION. The world's first virtual nation. Available at `https://github.com/Bit-Nation`, 08 2022.

[111] The World If. Disrupting the trust business. Available at `https://www.economist.com/the-world-if/2017/07/15/disrupting-the-trust-business`, 07 2017.

[112] Staci Warden. Can bitcoin technology solve the migrant crisis. Available at `https://www.wsj.com/articles/can-bitcoin-technology-solve-the-migrant-crisis-1465395474`, 06 2016.

[113] Jolocom. Available at `https://jolocom.io/`, 08 2022.

[114] Uniqueid. Available at `https://uniquid.com/`, 08 2022.

[115] Cryptid coin. Available at `https://cryptidcoin.io/`, 08 2022.

[116] Pinyaphat Tasatanattakool and Chian Techapanupreeda. Blockchain: Challenges and applications. *2018 International Conference on Information Networking (ICOIN)*, 01 2019.

[117] Marco Mazzoni, Antonio Corradi, and Vincezo Nicola. Performance evaluation of permissioned blockchains for financial applications: The consensys quorum case study. *Blockchain: Research and Applications*, 3(1), 2022. ISSN 2096-7209. doi: https://doi.org/10.1016/j.bcra.2021.100026.

[118] Yang Liu and Kim-Kwang Choo. Blockchain-based identity management systems: A review. *Journal of Network and Computer Applications*, 166, 09 2020.

[119] Kumaresan Mudliar and Harshal Parekh. A comprehensive integration of national identity with blockchain technology. *2018 International Conference on Communication information and Computing Technology (ICCICT)*, 02 2018.

[120] Nutthakorn Chalaemwongwan and Werasak Kurutach. A practical national digital id framework on blockchain (nidbc). *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 01 2018.

[121] Republic of Estonia. The new digital nation. Available at https://www.e-resident.gov.ee/, 08 2022.

[122] Becky Buckle. The home office is trialling a new "digital nightlife id". Available at https://mixmag.net/read/the-uk-are-trialling-a-new-digital-id-news/, 05 2022.

[123] 1account. About 1account. Available at https://www.1account.net/, 05 2022.

[124] PASS. The national proof of age standards scheme. Available at https://www.pass-scheme.org.uk/, 05 2022.

[125] PASS. Consultation on pass proposal to develop uk standards for the physical presentation of digital proof of age (dpoa). Available at https://www.pass-scheme.org.uk/wp-content/uploads/2020/10/Summary-of-responses-to-PASS-Consultation-1.pdf, 10 2020.

[126] PASS. Requirements for digital presentation of proof of age. Available at https://www.pass-scheme.org.uk/wp-content/uploads/2022/05/PASS-5-2021-Requirements-for-Digital-Presentation-of-Proof-of-Age.pdf, 2021.

[127] Abi Tunggal. The 64 biggest data breaches (updated may 2022). Available at https://www.upguard.com/blog/biggest-data-breaches, 2022.

[128] Apple. iphone 11 technical specifications. Available at https://www.apple.com/uk/iphone-11/specs/, 08 2022.

[129] Eugenia Politou, Fran Casino, Efthymios Alepis, and Constantinos Patsakis. Blockchain mutability: Challenges and proposed solutions. *IEEE Transactions on Emerging Topics in Computing*, PP:1–1, 10 2019. doi: 10.1109/TETC.2019.2949510.

[130] Morris Dworkin. Sha-3 standard: Permutation-based hash and extendable-output functions. 2015.

[131] C Shaw. United kingdom population trends in the 21st century. *Population trends*, 103:37–46, 02 2001.

[132] Gary Knott. Hashing functions. *Comput. J.*, 18:265–278, 03 1975. doi: 10.1093/comjnl/18.3.265.

[133] Cremers Cas. The scyther tool. Available at https://people.cispa.io/cas.cremers/scyther/, 04 2014.

[134] Yutaka Sasaki. The truth of the f-measure. *Teach Tutor Mater*, 01 2007.

[135] Daniel Barragán and Byung Eu. Irreversible thermodynamics of neural networks: Calortropy production in logic operations. *Journal of Physical Chemistry B - J PHYS CHEM B*, 105, 06 2001. doi: 10.1021/jp004504v.

[136] Andrey Zhmoginov and Mark Sandler. Inverting face embeddings with convolutional neural networks. 06 2016.

[137] Qiong Cao, Li Shen, Weidi Xie, Omkar Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. 10 2017.

[138] Tensorflow. Model zoo. Available at https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf2_detection_zoo.md (2022/08/19).

[139] Mingxing Tan, Ruoming Pang, and Quoc Le. Efficientdet: Scalable and efficient object detection. pages 10778–10787, 06 2020. doi: 10.1109/CVPR42600.2020.01079.

[140] Silvia Rostianingsih, Alexander Setiawan, and Christopher Halim. Coco (creating common object in context) dataset for chemistry apparatus. *Procedia Computer Science*, 171:2445–2452, 01 2020. doi: 10.1016/j.procs.2020.04.264.

[141] Heartex Labs. Labelimg. Available at `https://github.com/heartexlabs/labelImg`, 08 2022.

[142] Usha Ruby and Vamsidhar Yendapalli. Binary cross entropy with deep learning technique for image classification. *International Journal of Advanced Trends in Computer Science and Engineering*, 9, 10 2020. doi: 10.30534/ijatcse/2020/175942020.

[143] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014.

[144] Shabana Habib and Noreen Khan. An optimized approach to vehicle-type classification using a convolutional neural network. *Computers, Materials and Continua*, 69:3321–3335, 08 2021. doi: 10.32604/cmc.2021.015504.

[145] Solveig Badillo, Balazs Banfai, Fabian Birzele, Iakov Davydov, Lucy Hutchinson, Tony Kam-Thong, Juliane Siebourg-Polster, Bernhard Steiert, and Jitao David Zhang. An introduction to machine learning. *Clinical Pharmacology Therapeutics*, 107, 03 2020. doi: 10.1002/cpt.1796.

[146] Haider Allamy. Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). 12 2014.

[147] Ethereum Foundation. Go ethereum. Available at `https://geth.ethereum.org/`, 08 2022.

[148] Ethereum Foundation. Sepolia. Available at `https://sepolia.dev/#`, 08 2022.

[149] Etherscan. Gas tracker. Available at `https://etherscan.io/gastracker`, 08 2022.

[150] Coinbase. Ethereum (eth). Available at `https://www.coinbase.com/price/ethereum`, 08 2022.

[151] OpenJS. Node.js. Available at `https://nodejs.org/en/`, 08 2022.

[152] OpenJS. Express. Available at `https://expressjs.com/`, 08 2022.

[153] Web3.js. Available at `https://web3js.readthedocs.io/en/v1.7.5/`, 08 2022.

[154] Amazon. Aws ec2. Available at `https://aws.amazon.com/ec2/`, 08 2022.

[155] Facebook. React native. Available at `https://reactnative.dev/`, 08 2022.

[156] MongoDB. Realm react native. Available at `https://www.mongodb.com/docs/realm/sdk/react-native/`, 08 2022.

[157] React native keystore. Available at `https://www.npmjs.com/package/react-native-secure-key-store`, 08 2022.

[158] Google. Nearby messages api. Available at `https://developers.google.com/nearby/messages/overview`, 08 2022.

[159] Truffle. Truffle suite. Available at `https://trufflesuite.com/`, 08 2022.

[160] Gitlab. Gitlab ci/cd. Available at `https://docs.gitlab.com/ee/ci/`, 08 2022.

[161] Anamika Chauhan, Om Malviya, Madhav Verma, and Tejinder Singh Mor. Blockchain and scalability. pages 122–128, 07 2018. doi: 10.1109/QRS-C.2018.00034.

Appendix A

User Guide

A.1 Installation

The digital proof-of-age app that demonstrates the functionality of the identity management system is available for download in apk format **here**. This can be used on an android device with developer mode enabled. A guide to enable developer mode can be found **here**.

If an android device is not available then an emulator can be set up using android studio (note that verifications cannot be trialed with this solution, as local data transfer is not possible). A detailed guide to configuring an emulator can be found **here**. To install the apk on the emulator, drag and drop the file from the system file explorer onto the device screen.

The smart contract is deployed to Sepolia [148] and the governing server is live on an AWS instance so no further setup is required to test the software.

To test verifications and account transfer, two physical devices are required. If these are not available, see the walk-throughs below that illustrate the application's function.

A.2 Walk-through

A.2.1 Sign-up

The sign-up process in the application consists of a series of forms to collect the user's information, shown in Fig. A.1. Upon completion, the user is confronted with a page instructing them to wait for authentication. This process should be completed within 15 minutes. When a certification has been issued, the app will display a digital representation of the user's identity information.

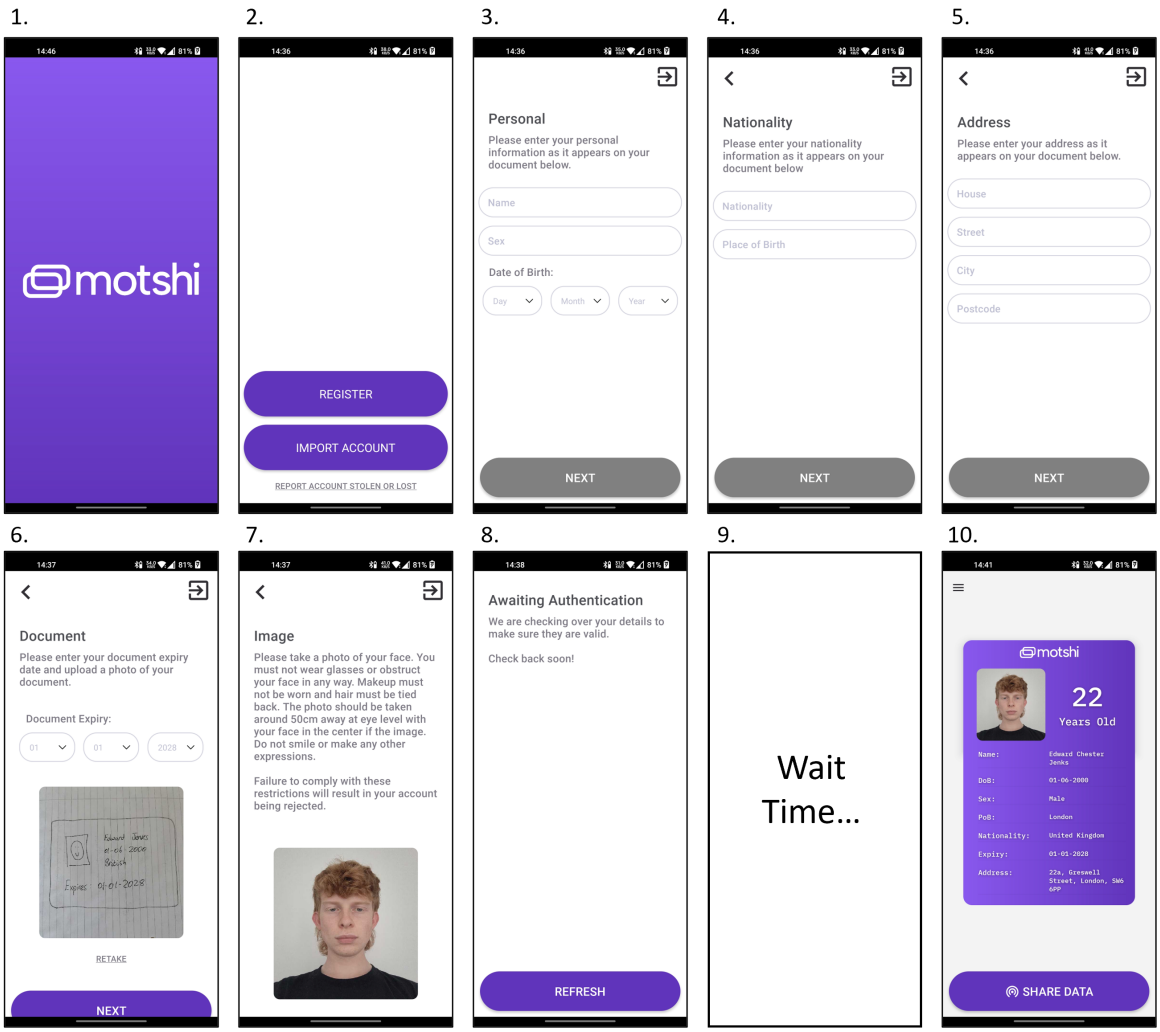

Figure A.1: Step-by-step sign-up process in digital proof-of-age app.

A.2.2 Verification

The verification process is shown in Fig. A.2. A verifier can navigate to the verification page from the navigation drawer. This will begin scanning for nearby users sharing their data. From the profile page, the user can choose to share their data. This will present a QR code on their screen. When the verifier's device detects the data transfer, a camera to scan the QR code will come up. After scanning, the verifier's device can check the user reported data matches the certification on the blockchain and presents the result of the verification. The user's data is then presented for inspection.

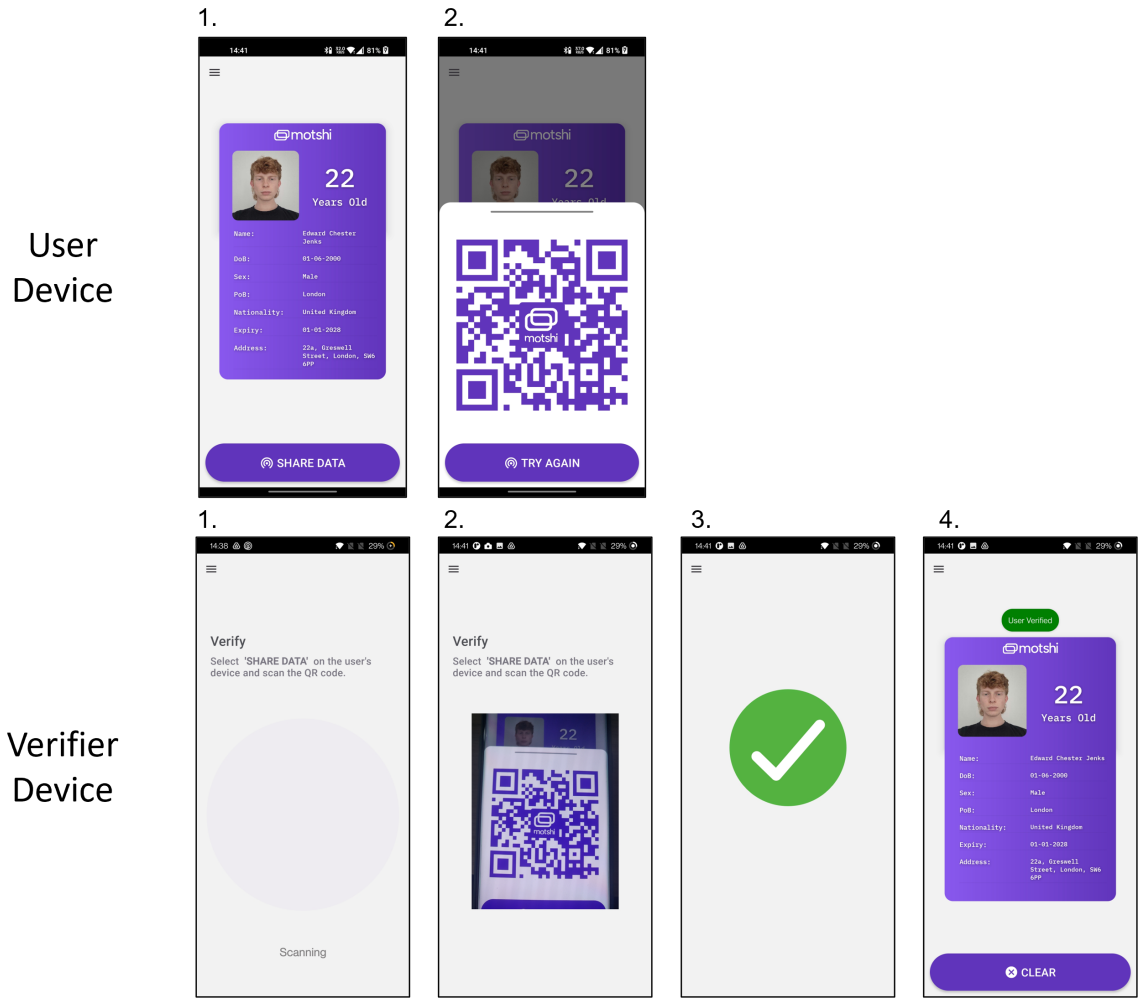

Figure A.2: Step-by-step verification process in digital proof-of-age app.

A.2.3 Account Migration

The account transfer process is shown in Fig. A.3. On the new device, the option to import an account can be selected. On the old device, the option to move account can be selected from the settings menu. This begins a local data transfer of the user information. When this is detected on the new device, a QR code is presented for the old device to scan. On scanning, the certification is transferred to the new address and the old device returns to the sign-up screen while the new device displays the account information.

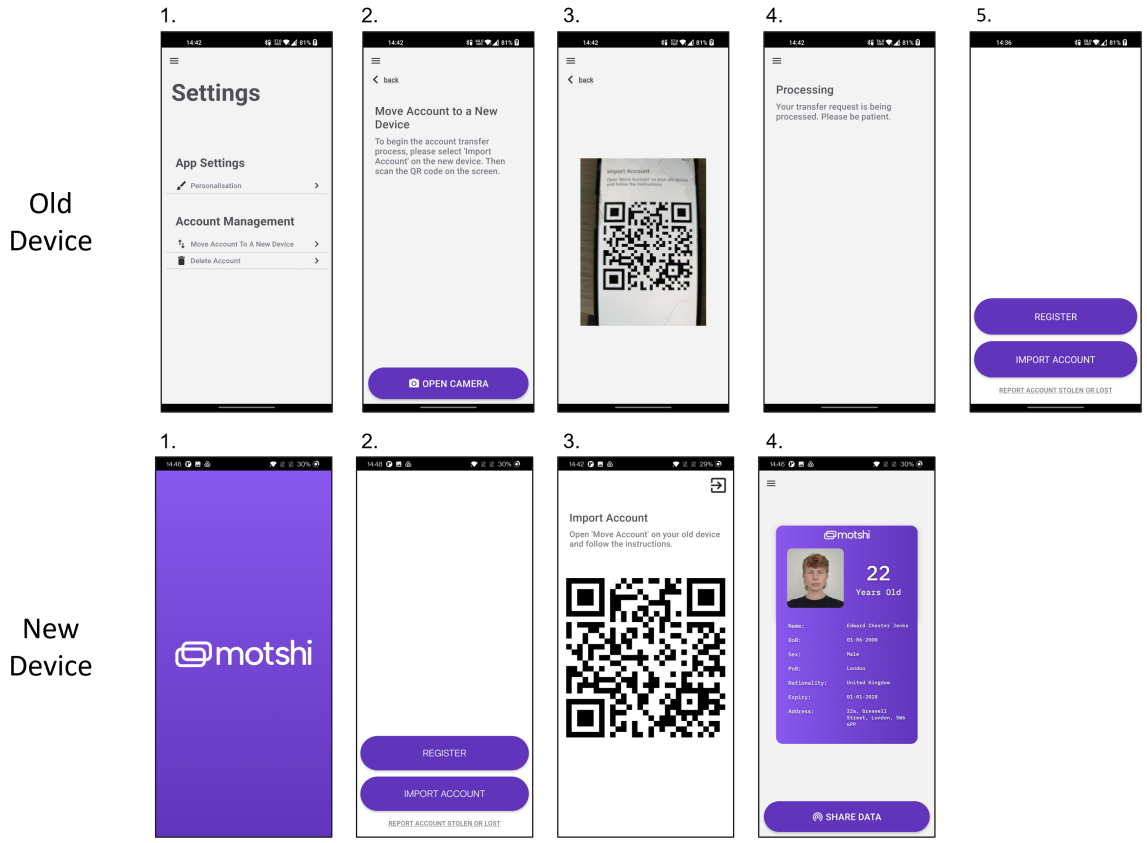

Figure A.3: Step-by-step account transfer process in digital proof-of-age app.