Cybersecurity Optimization and Training for Enhanced Resilience in Finance

D3.5 – Blockchain Security Focus whitepaper (I)

[WP3 – Cybersecurity Improvement in Digital Onboarding]

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833923. The contents of this publication are the sole responsibility of the authors and can in no way be taken to reflect the views of the European Commission.

Lead Contributor	Jose Manuel Panizo Plaza, FNMT-RCM (FNMT)
	josemanuel.panizo@fnmt.es
Other Contributors	Robin Renwick, Trilateral Research Ltd (TRI IE)
	Tina Ehrke-Rabel, Graz University (GRAZ)
	Roberto Fernández Hergueta (EVR)

Due Date	31.01.2020
Delivery Date	31.01.2020
Type	Report
Dissemination level	PU = Public

Keywords	SOTER, blockchain, security, cybersecurity

Document History

Version	Date	Description	Reason for Change	Distribution
V01.r01	02.12.2019	Draft	Initial draft	02.12.2019
V01.r02	22.12.2019	Revision	Amendments/Contributions	22.12.2019
V01.r03	23.12.2019	Revision	Amendments/Contributions	23.12.2019
V01.r04	14.01.2020	Revision	Amendments/Contributions	14.01.2020
V01.r05	15.01.2020	Revision	Amendments/Contributions	15.01.2020
V01.r06	24.01.2020	Revision	Amendments/Contributions	24.01.2020
V01.r07	26.01.2020	Revision	Amendments/Contributions	26.01.2020
V01.r08	29.01.2020	Revision	Amendments/Contributions	29.01.2020
V01.r09	30.01.2020	Revision	Amendments/Contributions	30.01.2020
V02	31.01.2020	Final	V02	31.01.2020

Abstract

This deliverable represents the first iteration of the SOTER D3.5 - Blockchain Security Focus whitepaper [M7]. The document provides information regarding the proposed blockchain implementation of the SOTER platform. It introduces a high-level overview of blockchain systems, consensus mechanisms, as well as specific design and architecture characteristics of blockchain systems. The deliverable provides a description of specific characteristics as they relate to the SOTER platform, as well as providing initial recommendations for the SOTER project based on information available as per the deliverable date (31st January 2020). The document outlines specific choices that require further consideration by the consortium, especially considering legal and regulatory compliance and the development and integration of specific protocols such as Self-Sovereign Identity, Decentralized Identifiers, and eIDAS, These considerations are relevant to the proposed Digital Onboarding Platform, and the role that digital identity plays within it, impacting aspects such as Data Protection, Privacy, GDPR compliance. The document also provides information regarding related tasks, with relevant deliverables viewed as important related reading, as they relate to security, risk, security audits, privacy impact assessments, as well as the Security by Design and Privacy by Design methodologies the SOTER project and consortium has agreed to adhere to, as outlined in the Grant Agreement. This initial document is seen as the first iteration, with the second due to be delivered at a later stage of the development lifecycle [M19].

Table of Contents

Executive Summary

SOTER is a European Commission H2020 funded project, entitled *'CyberSecurity Optimization and Training for Enhanced Resilience in finance'* (SOTER), Grant Agreement No. 833923. This deliverable is part of the work package entitled: *'Cybersecurity improvement in Digital Onboarding'*. The deliverable details security related aspects of Distributed Ledger Technology (DLT), otherwise known as Blockchain Technology.

The purpose of this document is to outline security concepts relating specifically to Blockchain Technology, considering concerns related to the SOTER project, its proposed development plan, and also to outline discussion related to specific choices that require to be made by the consortium, along with the justification for any choices, or recommendations that have been made. In summary, this document:

- Provides an introduction to the SOTER project, including the technical development aspects of the Digital Onboarding Platform that relate to blockchain technology
- Details a high-level overview of security issues related to the implementation of blockchain technology
- Provides security focused recommendations for the SOTER project with respect to specific implementations of blockchain technology
- Describes aspects of the technical implementation chosen for the SOTER project, along with justification regarding any recommendations

This document proposed a layered approach to blockchain technology security, incorporating broad analysis of the security model; business layer, governance layer, aspects of data access and network membership, consensus mechanisms, and identification processes, GDPR considerations, as well as specific security concerns regarding the data, application and infrastructure layers.

Following on from this, the document provides recommendations for the SOTER project, including recommendations concerning the business, governance, data, application, and infrastructure layers. These may be summarised as:

- The most suitable architecture for the SOTER project at this stage of development seems to be a private and permissioned blockchain implementation.
- The most suitable blockchain platform is HyperLedger Besu, which the ALASTRIA network is based on
- Decisions regarding Self-Sovereign Identity, Decentalized Identifiers, and eIDAS need to be appraised and considered by the consortium ongoing development of the European Blockchain Services Infrastructure
- Aspects of the ALASTRIA_ID solution should be considered for the SOTER project, with consideration given to GDPR and 5AMLD compliance as well as proposed interoperability with eIDAS

The deliverable provides information regarding related reading and relevant associated deliverables, especially those that consider security aspects of the SOTER platform along with those that impact on both Privacy by Design and Security by Design methodologies. These related deliverables are summarised as belonging to the following SOTER research tasks:

- T2.2 – Privacy Impact Assessment and Privacy by Design
- T3.2 – System Risk Assessment
- T3.5 – Potential Security Issues and Responding
- T3.6 – Auditing the Platform

This document notes that the document currently being read is the first iteration, delivered at M7, with an updated iteration due in M19.

List of Tables

No table of figures entries found.

List of Figures

List of acronyms/abbreviations

Abbreviation	Explanation
DLT	Distributed Ledger Technologies
SOTER	Cybersecurity Optimization and Training for Enhanced Resilience in Finance
DoA	Description of Action
DLT	Distributed Ledger Technology
IT	Information Technology
PoW	Proof-of-Work
PoS	Proof-of-Stake
WP	Work Package
BFT	Byzantine Fault Tolerance systems
FBA	Federated Byzantine Agreement
PoI	Proof-of-Importance
PoA	Proof-of-Authority
PBFT	Practical Byzantine Fault Tolerance
PKI	Public Key Infrastructure
HSM	Hardware Security Modules
NIST	National Institute of Standards and Technology

Table 1 List of acronyms/abbreviations

1.-Introduction

This deliverable is included in the SOTER work package (WP3) entitled "Cybersecurity improvement in Digital Onboarding". It is the first iteration of the document entitled "Blockchain Security Focus whitepaper". The deliverable replies to the requirements set in the Description of Action (DoA):

> "Due to Block Chain technology is in place, this analysis will focus on those aspects that are critical for this technology. Without them, the chain can be corrupted or become untrusted, or the chain data is accessed by unauthorised parties. 'Is this way of using block chain the most secure?', 'who is authorised to add a new block?', 'who is authorised to access to the chain data?' , 'Is the chain always available to add a new block?', 'how we know that the block to add is not a rouge one?', 'how we ensure that the chain is not a rouge one?'; They are some examples of questions we have to answer to ensure security in block chain technologies. Taking advantage of this exercise, we could produce a whitepaper with the main aspects to think about in terms of security for these cutting-edge technologies, such as the aspects mentioned in the first task."[1]

This report is based on information that exists at the deliverable date – 31st January 2020 (M7).

1.1.- SOTER Research

SOTER will research and develop a technology platform coupled with a user training methodology and manual to enhance cybersecurity resilience within organisations. The holistic research and development approach predominantly targets the financial services sector. The project will develop a Digital Onboarding Platform (DOP) incorporating biometric identification and authentication technology. The tool leverages blockchain technology to improve security, accessibility, robustness, audibility, and verifiability. A set of training methodologies will be developed targeting proposed end-users of the DOP. The training methodology will be based on the research outcomes of a qualitative study that explores the human factors of cybersecurity risk and the interdisciplinary development of a cybersecurity competence catalogue.

[1] T3.3 Task Description, SOTER Grant Agreement, p. 22.

1.2.- Scope of this deliverable

The deliverable is the first iteration of the blockchain security whitepaper. It provides an introduction to distributed ledger technology, otherwise known as blockchain technology. The document provides an overview of the key security considerations related to the implementation of blockchain technology for the SOTER project, and more specifically the implementation related to the technical development of a biometric based identification and authentication Digital Onboarding Platform (DOP).

The object of this deliverable is to:

- Provide an introduction to the technical development aspects of the SOTER project
- Detail the primary security related aspects of blockchain technology implementations
- Offer an overview of the security related recommendations for the SOTER project
- Provide an overview of the blockchain implementation chosen for the SOTER project, along with justification and rationale for the choices that have been taken

1.3.- Structure of this deliverable

This deliverable contains three main sections, along with an introduction and a conclusion. Following the introduction, Section 2 provides an outline to the key security considerations of blockchain technology and details a layered approach to blockchain technology security analysis. Section 3 considers each of the layers in specific relation to the SOTER project, outlining some of the key decisions required to be made regarding the technical development of the SOTER Digital Onboarding Platform (DOP). Section 4 provides information regarding decisions that have been taken at this initial stage of the SOTER project. This section also provides justification and rationale for the decisions. The final section provides a conclusion to the document.

1.4.- Relation to other deliverables

This document is part of the overarching *WP3*, entitled *'Cybersecurity Improvement in Digital Onboarding'.* The work package contains a number of related deliverables. The predominant task related to this deliverable is *T3.2 – System Risk Assessment*, and its related deliverables, the first iteration of which has been delivered in December, 2019 (M7). T3.2 seeks to provide a full risk assessment of the SOTER DOP. The first deliverable of this task has been delivered, and is entitled *D3.3 – System Risk Assessment*. The document provides an introduction to the risk assessment methodology, and an initial risk assessment of the platform in its current state of development. It is viewed as recommended reading for this document, which is seen as a more high-level architectural analysis of the SOTER platform implementation of blockchain technology.

There are also a number of other tasks within WP3 that should be viewed as related to the security analysis of the SOTER platform, and these are found within *T3.5 – Potential Security Issues and Responding*, and *T3.6 – Auditing the Platform*. Work has not yet begun on these tasks and is due to start in M12.

The final related task is found within *WP2 – General cybersecurity aspects. Human Factor as internal threat*. This WP contains T2.2 – Privacy Impact Assessment and Privacy-by-Design. This task contains the associated Privacy Impact Assessment (PIA+) deliverables (D2.3 and D2.4), which are viewed to inform the technical development process and ensure the development process adheres to a privacy-by-design methodology. These deliverables are viewed as related reading for this document. The first iteration of the PIA+ is due in M14. Activities for this task have begun, developing alongside the technical development process.

2.– Blockchain security analysis

Blockchain technology may be viewed as an umbrella term, used to describe different information storage alternatives. This section begins by defining some key concepts on which the definition of blockchain technology is based. After that, an attempt will be made to identify the risks that blockchains technologies present. The risks will be categorized through a top to bottom layered approach, which includes Business, Governance, Data, Application and Infrastructure layers.

2.1- Blockchain building blocks

The concept of a blockchain falls under a wider term which is called "Distributed Ledger Technologies", or DLT. This refers to the new paradigm in sharing and storing information. As the name suggest, it implies the existence of a:

- **Distributed database**: Data is not kept as a single copy, but rather it is stored distributed between nodes of a network. It can be seen as a database located in multiple places and being processed as a single unit. A distributed system should be consistent – each node has the same information at a certain point of time- and be failure tolerant, meaning that if one node fails to operate correctly this does not impact the correct functioning of the network as a whole.

- **Ledger**: The information is presented through a record of transactions between accounts or users of the infrastructure. These transactions can refer to the transfer of assets or to the change in the status of some piece of information, registered, shared, synchronized, and verified by the nodes of the network.

The above permits data to be stored resiliently, since it is kept by several nodes in a verifiable and transparent manner, due to the availability of the information.

A DLT may or may not be decentralized. Decentralization implies that the need for verification or approval by a central authority is removed, with responsibilities shared between different actors within the network. With such a property, data is filed independently from the validation of a trusted third party. The endorsement of the information is obtained through a consensus mechanism between some, or all, of the involved agents.

A blockchain is a type of DLT. It contains information in records, stored in blocks. The blocks are linked in a list using cryptographic techniques. Each block has some meta-information, such as a timestamp and the hash of the data contained in the previous block. When a block is appended to the chain, all the nodes of the network must reach agreement. If the block is confirmed, all the nodes must process it along with the included transactions. Although the most known use of blockchain technologies relates to cryptocurrencies, transactions should not be seen as only financial operations. A transaction is a just a change in the state of an item.

Blockchain technologies offers appealing characteristics such as:

- **Immutability**: Once data is stored in a block, and the block is confirmed by the network, the information cannot be altered without rebuilding the remaining chain. This is a layer of security which ensures that data, once appended to the chain, cannot be altered by any one party without substantial effort and significant resources. Any change in previously confirmed data must be confirmed, and accepted, by other members in the network through the protocol consensus mechanism.

- **Transparency**: Transactions and changes of state of data are shared between all the organizations that have permission to view the stored information. This adds a degree of accountability to the data store, which is extremely beneficial for the financial industry, as state changes are recorded accurately, and consistently, according to the protocol rules of the network.

- **Trustless**: Participants in the network agree to run a consensus protocol, used to reach agreement on new state changes and/or transactions. There is no one source of truth for the network, as the order of events is agreed upon by all parties in a synchronous manner. In this way, a decentralized ledger is a trustless system as there is no need to delegate trust to a third-party for an agreed order of events within the network.

2.2.- Blockchain Security Model

Once some basic concepts are defined, the next step is to establish a framework which will allow the SOTER consortium to evaluate issues that impact the security of the chosen blockchain implementation. This is viewed as a starting point to address risks that should be considered in order to ensure the blockchain is aligned with the security restrictions of the extended finance sector.

In order to identify the risks related to a blockchain implementation, an analysis is proposed following the model suggested by Arunkumar and Muppidi [2]. It is set as a layered model which approaches the risks related to any deployed blockchain solution. The following layers are addressed:

2.2.1.- Business Layer

As in any technological project, the first step towards achieving success is to guarantee that Information Technology (IT), security and business strategies are aligned and coherent. It must be ensured that the use of blockchain is relevant for the project, and if the business requirements are clear and fulfilled by this technology. Blockchain should be useful enough to add value to every process of the project. The flow diagram presented in *Blockchain Beyond the Hype*[3] will support this analysis, as it develops a set of questions used to ascertain whether blockchain is required for a business. Firstly, it presents three compulsory requirements:

- The necessity to remove intermediaries or brokers.
- The need to deal with digital assets.
- The requirement to create a permanent authoritative record of the digital assets.

It also reflects on two characteristics that are not yet met by blockchain implementations in an efficient way:

- High 'performance' while processing transactions.
- A large amount of non-transactional data stored in the solution.

The document also affirms the need to develop some further research if the business requires the following objectives:

- There exists a reliance on a centralised trusted party.

[2] Saritha Arunkumar and Sridhar Muppidi (2019). Secure your blockchain solutions. IBM. Retrieved from https://developer.ibm.com/articles/how-to-secure-blockchain-solutions/

[3] Rangaswami, J., Warren, S., Mulligan, C., & Zhu Scott, J. (2018). Blockchain Beyond the Hype A Practical Framework for Business Leaders. White Paper in World Economic Forum 2018, (April). Retrieved from http://www3.weforum.org/docs/48423_Whether_Blockchain_WP.pdf

- Shared write access to partial data stored in the blockchain is required.

If a trusted party is needed, or shared write access is not required, it is viewed as unnecessary to deploy a blockchain.

The document also states that these two requirements contribute to provide true value to a blockchain implementation.

- The solution is going to reflect contractual relationships and value exchange of digital assets.
- The actors/entities do not know or trust each other or may have misaligned interests.

In the next section, these questions will be considered for the SOTER project.

2.2.2.- Governance Layer

The organisational model of the blockchain network is as important as the underlying technology, so a governance framework should be established to create guidelines and procedures which define:

- Who is granted read and write access permission to the information stored in the blockchain.
- How participants in the blockchain reach consensus on any proposed appending of data.
- How members are identified, and how their credentials are managed and/or appraised.
- Within the scope of GDPR, what is the role of each member.

In the next sub-sections, these issues will be discussed to reflect all the considerations that should be taken into account to assess the security of a blockchain.

2.2.2.1.- Data access and network membership

In this section, a taxonomy will be established to classify the different blockchain implementations that are available currently. The classification will be based on categories

proposed by BitFury Group[4], clarified in the following sub-section. In the next chapter it will be analysed which should be the proposed implementation for the SOTER project.

2.2.2.1.1.- Data Access: Public and Private blockchains

From the point of view of establishing access and modification to the stored information, a blockchain can be defined as public or private. In a public blockchain, there are no restrictions on reading data or on who can propose transactions to be included by the network (append/modify data). It is generally riskier, as anyone can take part in the blockchain mechanism, as long as they partake according to the consensus ruleset.

In a private blockchain read access and append/modify access is limited to a closed group of nodes, which are controlled by a regulator or private consortium. The nodes in the network all agree to partake according to the consensus ruleset.

2.2.2.1.2.- Network Access: Permissioned and Permissionless blockchains

Taking into account the perspective of which nodes can process transactions, other classification can be established. In a permissionless blockchain, the capacity of joining the network is offered to any prospective node. There are no rules for joining the network and for becoming involved in the processing of transactions.

By contrast, in a permissioned blockchain, only a close list of identified and vetted nodes are able to join the network and process transactions. This divergence is related to the governance of the blockchain, and it should be considered when the recommendation for the selected implementation is formulated for the SOTER project.

2.2.2.2.- Consensus mechanisms

Due to the distributed property of a blockchain network, it is required that an agreement is reached between nodes on the objective version of events. The consensus mechanism and ruleset are required to define the rules deployed to validate new information. This agreement

[4] Garzik, J. (2015). Public versus Private Blockchains Part 1: Permissioned Blockchains White Paper, BitFury Group in collaboration with. Retrieved from https://bitfury.com/content/downloads/public-vs-private-pt1-1.pdf

is called a "consensus mechanism" and can be executed by all the nodes in the network, or a portion of nodes in the network. Its goal is to determine if a transaction is valid or not, using cryptographic validation methods. In addition, the agreement is needed to resolve any conflicts that arise if an entity attempts to append false or fraudulent information to the database.

In the next sub-sections, some of the most popular consensus mechanisms will be explained. This is relevant to which is most applicable to the SOTER project requirements.

2.2.2.2.1.- Proof-of-Work

In a Proof-of-Work (PoW) mechanism 'a prover demonstrates to a verifier that she has performed a certain amount of computational work in a specified interval of time'[5] . Here, a previous relationship of trust is not needed between nodes or with a third-party. It is based on the computing power of participant nodes. A cryptographic challenge is set by the consensus protocol, which can be viewed as work required to be completed by a node and proof of completion provided. The first node which accomplishes it is the responsible for appending a block to the chain. This block contains some of the pending transactions. In some cases, the appending node earns some reward for performing this operation. In addition to the appending node is responsible for adding the block to the chain and informing the rest of the nodes of the appended block.

This mechanism is resistant to a fraudulent attack, a situation where one of the nodes sends a transaction with false information expecting the network commit and confirm this block. In this case, a bifurcation of the chain will occur, one side with legitimate information and the other side with the fraudulent data. If the honest nodes in the network are in the majority (>50%), the legitimate side of the chain will be accepted by the majority of the network. For this reason, this is called "51 % attack". A blockchain is only resistant to this attack if the majority of the network are honest. A more detail reading about this fraudulent attack resistance can be found in the article entitled by Rijnbout, 2017.[6]

[5] Jakobsson, M., & Juels, A. (1999). Proofs of Work and Bread Pudding Protocols (Extended Abstract). In Secure Information Networks (pp. 258–272). Retrieved from https://doi.org/10.1007/978-0-387-35568-9_18
[6] Rijnbout, J. (2017). Byzantine Consensus Through Bitcoin's Proof-of-Work. In Management Control & Accounting.

2.2.2.2.2.- Proof-of-Stake

The term *Proof-of-Stake* (PoS) refers to a set of consensus mechanisms that depend on validator's economic stake in the network[7]. They are used in public blockchains, such as Ethereum, which is switching its blockchain from PoW to a PoS system.[8] In PoS, miners are replaced with validators. They take turns to propose and vote on the next block. The vote system is weighted, and the weight the vote of each validator depends on the validator's deposit or stake. PoS blockchains uphold an internal base cryptocurrency which reflects the interest of a validator on the network and its rights to participate in the consensus mechanism.

In other words, it is established that the node who owns more assets of the distributed network, will have a higher chance of being provided the ability to append a block to the chain. The overarching principle is that the more assets a node has, the more incentive it has to act honestly (or at least in the best interests of the network). In this mechanism, the node must also provide prove that it has access to the mentioned assets.

Comparing this mechanism to PoW, it has several advantages. PoS is more efficient due to the lack of computation effort required by the protocol. The scalability of the network is improved because the average the time taken to propose the block-creator is shortened, which makes the consensus mechanism more efficient. However, a fraudulent attack might be easier because it does not need a high resource effort (electricity and computation power), so the implementation of a network using this agreement should also consider additional security measures. For instance, some networks have the possibility of blocking assets that a node owns and the ability to requisition them if fraud is committed. In the Ethereum blockchain, validators join by sending an amount of the internal cryptocurrency to a special smart contract. They are kept in a 'lock box', which is only changeable if the validator acts

[7] Buterin, V. Proof of Stake FAQ. Retrieved January 26, 2019, from https://github.com/ethereum/wiki/wiki/Proof-of-Stake-FAQ#what-is-proof-of-stake

[8] Ethereum 2.0 Phases. (2019). Retrieved from https://docs.ethhub.io/ethereum-roadmap/ethereum-2.0/eth-2.0-phases/

honestly of maliciously. In the Ethereum official documentation[9] it is possible to find out more information about this mechanism.

2.2.2.2.3.- Threshold signature scheme

Threshold signature schemes [10] are another sort of consensus mechanism which is based on a multiple signature scheme. This method starts with the assumption that there are a set of candidate nodes to confirm new information, and the signature of a subset of them is needed to perform this action. To avoid a fraudulent node attack, it is compulsory that the number of required signatures should be higher than half of the total available nodes [1/2(N)+1]. In particular for a blockchain, following this agreement, it is required to confirm a block that it is signed, at least, by half plus one of the nodes. In addition, it is compulsory to have a reliable Public Key Infrastructure (PKI) to provide a layer of trust to the platform to ensure that fraudulent keys are not distributed within the network.

The latency provided by this mechanism is low, but it has the disadvantage that a Public Key Infrastructure is opposed to the decentralized goal of a blockchain network.

2.2.2.2.4.- Federated Byzantine Agreement

This consensus mechanism originated from the concept of Byzantine Fault Tolerance systems (BFT). These systems tolerate failures related to the Byzantine Generals' Problem[11], which describes a situation where a group of distributed computing systems communicate to reach

[9] Ryan, Danny, Diederik Loerakker, Carl Beekhuizen, Hsiao-Wei Wang, B. E. (n.d.). Ethereum 2.0 Phase 0 -- The Beacon Chain. Retrieved January 26, 2020, from https://github.com/ethereum/eth2.0-specs/blob/dev/specs/phase0/beacon-chain.md

[10] Wall, E., & Malm, G. (2016). Using Blockchain Technology and Smart Contracts to Create a Distributed Securities Depository. In Lund University. Retrieved from http://www.eit.lth.se/sprapport.php?uid=987

[11] Lamport, L., Shostak, R., & Pease, M. (1982). The Byzantine Generals Problem. In ACM Transactions on Programming Languages and Systems (TOPLAS) (Vol. 4). https://doi.org/10.1145/357172.357176

agreement while some of them are corrupt and disseminate false information. A BFT group of nodes is able to continue its operation even if a sub-group of them is acting maliciously. Within a Federated Byzantine Agreement[12] (FBA) system, the mechanism is limited to a closed list of nodes. In other words, it is restricted to a permissioned network. The was one of the first methods used in blockchain networks. The nodes validate transactions blocks when they reach an agreement, which is accomplished when a minimum number of nodes agree. This number, known as quorum, is determined by the protocol ruleset.

Each participant node chooses to trust a given subset of nodes. A predefined list of validators is not needed. Nodes decide individually who to trust. Inside this close circle local consensus is reached. When a node trusts in another node which belongs to other circle, this local consensus will be spread out over the network. Therefore, it is needed that all the trust circles have common nodes and that they are not disjoint sets. If the circles are disjoint, agreement in the network can be blocked. This is the risk of FBA systems. A detailed explanation of FBA voting systems can be read in the Stellar Protocol blog[13].

2.2.2.2.5.- Practical Byzantine Fault Tolerance

Practical Byzantine Fault Tolerance (PBFT)[14] provides a different mechanism for reaching consensus. It is based on the FBA method, but with a slight variation. Every consensus round, one node is chosen as a leader in a round-robin agreement, and the rest of them act as systemic backup. Each node implements a state machine to develop a consensus algorithm in four steps. The process starts when a client sends a transaction request to the leader. The request is sent to the backup nodes, which execute the request and send the reply to the client. When the client receives a certain number of identical replies, it can accept the reply and be assured that consensus has been reached. The messages are signed to provide

[12] Mazi`eres, D., & Mazi`eres, M. (n.d.). The Stellar Consensus Protocol: A Federated Model for Internet-level Consensus. Retrieved from https://www.stellar.org/papers/stellar-consensus-protocol.pdf

[13] Foundation, S. D. (n.d.). On Worldwide Consensus – A Stellar Journey – Medium. Retrieved January 26, 2020, from https://medium.com/a-stellar-journey/on-worldwide-consensus-359e9eb3e949

[14] Castro, M. (2001). Practical Byzantine Fault Tolerance. In Proceedings of the Third Symposium on Operating Systems Design OSDI '99. Retrieved from http://pmg.csail.mit.edu/papers/osdi99.pdf

authenticity and integrity to the communication. As a result, a Public Key Infrastructure (PKI) is required.
This method is quicker than previous ones, so it can be applied to high-demand systems. The main disadvantage of this agreement is the fact that due to the high number of messages interchanged between nodes, it is not suitable for a wide network. The messages contain digital signatures and integrity codes, so there are also issues with practicality in a network with a high number of nodes.

2.2.2.2.6.- Proof of Importance

Proof-of-Importance (PoI) is a recently developed consensus mechanism. It is based on PoS protocol but is dissimilar, as the node selected to append a block is the one that has the highest 'score'. While this score is the number of assets owned by the node as in PoS, PoI more variables are considered as can be read in one implementation of it[15]. The calculation includes primary inputs such as the assets spent in the past 30 days, or the number of transactions executed in a given time period.

2.2.2.2.7.- Proof of Authority

Proof-of-Authority (PoA)[16] is a recently developed protocol based on reputation provided by the identities of the blockchain members. For this reason, its definition is in contrast with other consensus mechanisms such as PoW or PoS, which do not deal with real identities. This means that transparency is guaranteed by the reputation and the real identity of nodes.

The validation process begins from a closed list of identified nodes, whose role affords them the capability of validating blocks. The nodes are inserted into this list as a result of a voting

[15] NEM Technical Reference (2018). Retrieved from https://nem.io/wp-content/themes/nem/files/NEM_techRef.pdf

[16] Arasev, V. (n.d.). Proof-of-Authority Network Whitepaper. Retrieved January 26, 2020, from https://github.com/poanetwork/wiki/wiki/POA-Network-Whitepaper

system, completed by a set of authorized nodes. The nodes are chosen randomly to sign and append a block.

As this consensus mechanism does not contain high-computation cryptographic requirements, this consensus method requires less resources than others. Its strength relies on the fact that the nodes are pre-identified, and they need to protect this reputation. If they commit fraud their reputation will be affected, and so they have an inventive to continue to act in the best interests of the network. It should also be noted that due to the small number of messages sent, this mechanism offers a great degree of scalability and performance.

2.2.2.2.8.- Istanbul Byzantine Fault Tolerance

Istanbul Byzantine Fault Tolerance (IBFT)[17] is inspired by the implementation of a PoA protocol such as Clique[18]. The author of this mechanism was inspired by Practical Byzantine Fault Tolerance. They considered that there should not be clients that send messages and every node, called validators, can propose a block. In each consensus round, a verifiable new block is expected to be confirmed. The mechanism can tolerate a maximum of F faulty nodes in a network of comprised of 3F+1 nodes.

IBFT is similar to PBFT, in that a consensus round is based on three steps or phases. The round starts when the node selected as *Proposer* sends a new block proposal, along with a '*Pre-prepared message*'. When all the validators receive it, they broadcast a '*Prepare*' message. This step is needed to ensure that all nodes are synchronized (which means that they are working in the same round). When a validator has received 2F +1 '*Prepared*' messages, it broadcasts a '*Commit*' message. This means that the validator accepts the proposed block and is going to insert a block into the chain. Finally, each validator waits for 2F +1 '*Commit*' messages, and they insert the block to the chain. The entirety of the *Commit* messages received, are signed, and inserted as metadata into the proposed block.

---

[17] Yu-Te Lin. (2017). Istanbul Byzantine Fault Tolerance · Issue #650 · ethereum/EIPs · GitHub. Retrieved October 30, 2019, from https://github.com/ethereum/EIPs/issues/650

[18] Péter Szilágyi. (2017). Clique PoA protocol & Rinkeby PoA testnet · Issue #225 · ethereum/EIPs · GitHub. Retrieved October 30, 2019, from https://github.com/ethereum/EIPs/issues/225

The next figure, taken from[14], resumes the state transitions of the Istanbul Practical Byzantine Tolerance consensus protocol.

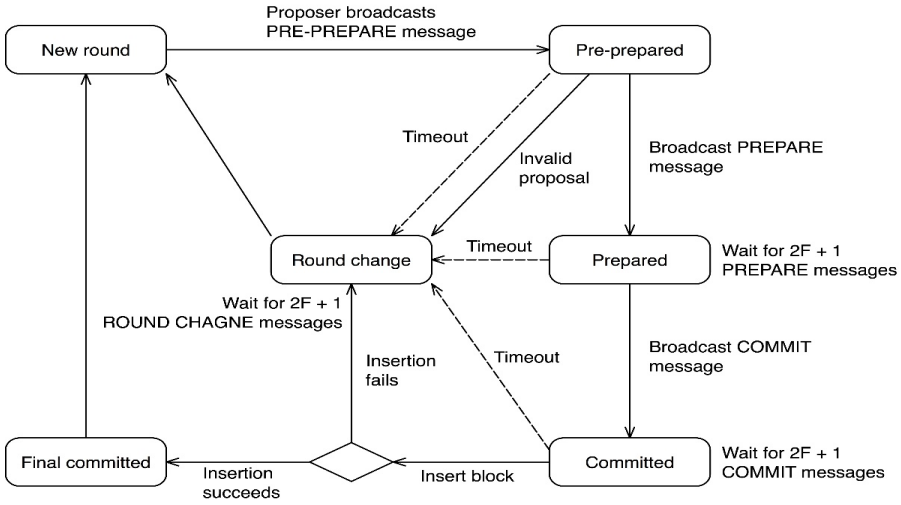

Figure 1. Istanbul Byzantine Fault Tolerance

A deeper analysis of this protocol can be found in the work of R. Saltini and D. Hyland-Wood[19].

2.2.2.2.9.- Raft

The Raft protocol[20] reaches consensus by choosing a single *Leader* node (I) from all the nodes in the network. This special node is assumed to always act honestly and is responsible for proposing a block and broadcasting it. The remaining nodes are *Followers,* and they just commit the proposals of the leader node. If the leader crashes, all the nodes are considered as *Candidates* to replace (I) as the *Leader*. In this situation, a new leader election starts. All the nodes maintain a common sequence, called term number. If a follower node does not receive any communication from the leader node after an arbitrary timeout, it is elected as a candidate. Then, the node increments the term number, votes itself, and sends a vote request

[19] Saltini, R., & Hyland-Wood, D. (2019). Correctness Analysis of IBFT. 1–31. Retrieved from http://arxiv.org/abs/1901.07160

[20] Diego Ongaro and John Ousterhout. (2015). In Search of an Understandable Consensus Algorithm. Retrieved from https://raft.github.io/raft.pdf

to all the nodes. If a node receives a request vote message with a term number higher than the registered value, it grants a vote to the sender. In normal conditions, the node that waits for the lower timeout will be selected as *Leader.*
Raft is a simple and quick crash fault tolerance algorithm which is able to recover from the failure of a node but does not provide Byzantine fault tolerance. Its use is only recommendable in a blockchain internally deployed which all nodes are well-known and controlled.

2.2.2.3.- Identification and authentication processes

One of the strengths of data storage using a blockchain is the requirement of consensus between all the involved nodes. In a permissioned network, it should be inferred that nodes should be properly identified and authenticated. These processes must be perfectly defined, implemented, executed and controlled. If not, a vulnerability can result in fatal fallouts such as unauthorized access, denial of service, information theft, or even loss. Besides that, a fail in these processes can cause considerable risks associated with compliance and/or litigation.

First, the governance body must determine the identification and authentication policies, depending of the type of blockchain developed. These policies should include:

- The list of entities which take part in the blockchain network (through the possession of a node), and how they will be identified.
- A secure storage mechanism where the set of authorized identities will be kept.
- The correct access levels that the entities own.
- The on-boarding mechanism to add a new member to the blockchain network.

Furthermore, a crucial aspect is the management of the credentials that the entities will use to authenticate the participant nodes in the blockchain network. These credentials should maintain a secure life cycle, and the processes for issuance, renewal, verification and revocation should be considered. Finally, the responsibilities of participants should be communicated through some form of contract when these credentials are activated.

A business-critical decision is whose responsibility it is to activate these credentials. This does not mean that they will be responsible for the Public Key Infrastructure (PKI) certificates and

nor will they be the Certification Authority, but at least, there should be an entity responsible to validate a member and to verify the claims made by them before activating their credentials. A mechanism to verify if the credentials are valid is also necessary, as well as a mechanism to activate them. The juxtaposition between the decentralized nature of blockchain technology and this sort of centralized responsibility of authorization oversight and revocation of member credentials should be noted.

These types of policies and roles will be discussed and developed during the SOTER project and reflected on within the second deliverable of this whitepaper (due M19).

Some initial recommendations for the SOTER project will be addressed in the following section.

As it is stated in the DoA of SOTER Project, one of the strengths of it is to empower end-users towards the use of their data. This means that they will have complete autonomy on the use of their personal information. This will be possible thanks to a sovereign identity. All the processes that support the identity system must be within the eIDAS regulation scope, which provides a powerful framework for digital identity and trust services. Due to this regulation, end-users will be able to use their credentials in all member countries. In addition to that, end-users of the SOTER project will only need to register its credentials once following the 'Once-Only' principle. These end-users' credentials will provide the identify attributes of eIDAS. They will also collaborate in the SOTER project capabilities to acquire verified data about end-users. Financial institutions will need valid data for use them in their administrative or business process. The veracity of this data must be clear, transparent and quick to establish. Manual labour should be avoided as much as possible.

Current identity management systems expose end-users' personal data to third party providers, and the user loses control of this data. They do not know who is using their personal information and the purpose for its use. The SOTER project will deploy a user-centric identity system, where, the user will become the sovereign owner of their information. The identity framework will also be GDPR compliant.

In the next chapter, a solution to achieve these requirements will be analysed, and we will see how blockchain technologies are incorporated.

2.2.2.4.- GDPR considerations for blockchain members

The blockchain network must assure compliance with data regulations including GDPR. This regulation includes the definition of different roles involved in data treatment. It defines "Data Controller" as:

> The natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law

This regulation also determines the role of "Data Processor" in this way:

> The natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller

Both roles own different responsibilities according to this regulation. In a blockchain, as it was stated in previous sections, a member can have read or write data access, but it can also update its local copy of the ledger or just execute the pieces of code included in a smart contract. Depending on the role assigned to each member, blockchain participants will be different responsibilities which must be informed in a contractual manner.

At a later stage in SOTER project this role will be discussed, and this section will be completed the second deliverable of the document in M19.

2.2.3.- Data Layer

The security aspects of the data layer are related to confidentiality, but also to data regulation and compliance.

By default, the information stored in the blockchain can be accessed by all network members. There are no privacy or confidentiality mechanisms in place. Therefore, if there is a need to store private data, security properties will be achieved by additional sub-layers. The mechanisms used to protect the information should be carefully analysed, because there might be a discrepancy between the desired confidentiality that the data storage should have, and the transparency offered by a blockchain.

Keeping in mind that blockchain technology offers data immutability, it is also necessary to examine which information is acceptable to be stored in this distributed ledger in terms of GDPR compliance and PSD2 privacy requirements. As a first approximation, there is a confrontation between GDPR and the immutability of blockchain[21]. This is a topic that requires careful consideration.

With regard to the data layer, some recommendations will be reflected in the next chapter, and a full discussion of this matter will be documented in the second iteration of this document (due M19)

2.2.4.- Application Layer

As with any infrastructure, the business logic encodes the business rules to interact with the stored data, for the purpose of solving the problems the application was created to solve. In a blockchain stack, at the application layer, a set of smart contracts encapsulate this logic to integrate real-world rules into the blockchain network. These pieces of software are self-executed in the nodes of the network and result in final transactions.

At this level, it is necessary to consider the following security issues concerning smart contracts:

- They are pieces of software, so they must follow the security procedures and best practices that are used in the software industry.
- They should be tested in a test environment, especially the blockchain implementations that ask for an execution fee (Ethereum).
- As any software, they can be vulnerable, so penetration testing to find security vulnerabilities and to identify security weakness is required.
- Depending on the consensus mechanisms implemented, there could be endorsement policies which establish the criteria to confirm the

[21] Blockchain and the General Data Protection Regulation, Can distributed ledgers be squared with European data protection law?, available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634445/EPRS_STU(2019)634445_EN.pdf

execution of a smart contract (for example, a number of member signature). All these requirements should be defined and can have implications at the governance layer.

Within the scope of GDPR, smart contracts need some serious thought regarding to automated processing. This regulation states in Article 22:

> "The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly affects him or her"

Throughout the course of SOTER project, this issue will be discussed, and this section will be completed the second deliverable of the document in M19.

To conclude, it must be taken into consideration that the on-boarding platform can require an interaction between client data and the blockchain network at any time. These blockchain clients, entitled *wallets*, are not in the scope of the SOTER project but if they are used, a minimal security posture should be defined for participation in the SOTER project.

2.2.5.- Infrastructure Layer

A blockchain network consists of nodes connected through communication networks. Nodes can be linked by an isolated network or using a set of complex networks connected by devices such as switches, firewalls or routers. In both cases, it is required to define and implement all the policies and practices needed to prevent, monitor and detect malicious actions such as unauthorized access, or denial of service.

At this layer, the following principles should be kept in mind:

- All the devices and nodes should be configured following best-practices guidelines and well-known security guides (provided by the manufacturer or security community).
- The attack surface of the blockchain network should be reduced as much as possible to limit the opportunities that an unauthorized user can have to extract information from the network.

- All the artefacts must be updated on time, following recommendations from the manufacturer. These update processes should be automatic, where possible, and should be monitored in the case of network or device error.
- Before the platform enters production and on a rolling basis in production, a vulnerability scan must be performed to find possible security weaknesses. This is relevant to the tasks included within the DoA, entitled *T3.2 – System Risk Assessment*, and more specifically *T3.6 – Auditing the Platform.*

3.- Recommendations for the SOTER project

In the previous section, a layered security model was presented to identify the risks and security considerations for a common blockchain implementation. Defining and identifying these considerations are necessary to specify some considerations and recommendations for a financial industry project like the SOTER DOP. In this section these requirements will be discussed. After that, the desired properties that the blockchain architecture will be established.

One of the key aims of the SOTER Project is to deploy a blockchain architecture to provide immutability, privacy and integrity of data but allowing for interoperability. The security that this technology is built upon should not be centralized. The blockchain must allow for data sharing between different application providers to enhance usability while safeguarding data privacy and security. Also, the blockchain infrastructure enables users to store credentials related to them on the network, allowing them to share and manage data between nodes. T

In terms of identity the user needs to register its credentials and this information, stored on the blockchain platform, can be shared among other parties, facilitating on-boarding and KYC (Know Your Costumer) processes. The objective is building a network where each node is a service/application from a different provider or company and the user can easily share her/his credential among the parties. Importantly, the user becomes the sovereign owner of this credential, and maintains control of the information that was used to earn this credential.

Taking into account certain requirements, establishing the characteristics that the chosen implementation of the blockchain should have will be discussed in this chapter and in more detail in the second iteration of the deliverable, due in M19.

The desired characteristics are presented in a layered model, as has been done in the previous section of this deliverable (see Section 2).

3.1.- Business Layer

In the previous chapter, a set of questions was presented to ascertain whether blockchain technology is sufficiently useful and adds value to the SOTER project. They were derived from the flow diagram presented in the whitepaper *Blockchain beyond the hype*[22]. In this section, these questions will be answered, leading to a discussion on the desired characteristics of a system designed for the extended finance sector, as SOTER is.

It is reasonable to suggest that the SOTER platform has three compulsory requirements for its chosen blockchain network. There is a requirement to remove intermediaries between trust providers, financial institutions and third parties.

SOTER project also deals with digital assets such as digital identities and/or KYC information. Besides that, it is possible to register a permanent authoritative record of these digital assets, applying cryptographic techniques such as digital signatures or hashing.

The SOTER project should consider that certain blockchain implementations cannot offer high performance of transaction throughput, usually measured in transactions per second (TPS). This is an issue that should be considered when a blockchain implementation is chosen. The throughput and performance of the network must be tested and discussed with stakeholders and financial and technical experts.

In terms of data storage, a blockchain is a specific type of database. It is not suitable for storing a large amount of non-transaction related data. If a high volume of information is required to

[22] Rangaswami, J., Warren, S., Mulligan, C., & Zhu Scott, J. (2018). Blockchain Beyond the Hype A Practical Framework for Business Leaders. White Paper in World Economic Forum 2018, (April). Retrieved from http://www3.weforum.org/docs/48423_Whether_Blockchain_WP.pdf

be kept, there could be a delay in read and/or write operations. An *off-chain* storage solution is recommended if a great volume of information is required to be stored, with a subsequent link between *off-chain* and *on-chain* information. This requirement should be considered in the design of the infrastructure.

Regarding the dependence on trusted parties, such as industry regulators, it is possible that the SOTER project will require their involvement to ensure legal and regulatory compliance. In this case, it will be necessary to develop an access point for them to have visibility of some aspects of the transaction data. This condition raises considerable security concerns surrounding privacy and/or confidentiality, because it is not desirable to display the entirety of the transaction data. Some cryptographic techniques, such as Zero-Knowledge Proofs[23] are applicable here. These protocols offer functions to prove that one possesses knowledge of certain information without revealing the actual information itself. Zero Knowledge Proofs are a secure, privacy preserving and privacy respecting cryptographic method for providing knowledge of a statement without revealing the contents of the statement to an external party.

In the instance of SOTER, shared write access to the blockchain will be required, especially in the case that the nodes may send multiple and/or almost simultaneous transactions.

Finally, the SOTER project reflects a contractual relationship between entities because it is possible to register consent, acceptance and verification of the on-boarding processes.

3.2.- Governance Layer

Taking into account the requirements of the SOTER Project, the characteristics of the chosen implementation of blockchain should have, with respect to the governance of the implementation will be discussed in this section.

[23] Introduction to Zero Knowledge Proofs: https://medium.com/@kotsbtechcdac/introduction-to-zero-knowledge-proof-the-protocol-of-next-generation-blockchain-305b2fc7f8e5

3.2.1.- Access to the stored information in the blockchain

The goal of this section is to define how information will be accessed, which nodes can process and/or propose transactions to the network (public or private), and the rules surrounding joining the network and being able to process transactions (permissioned or permissionless) in the blockchain solution for the SOTER project. For this analysis it is necessary to keep in mind the definitions explained in previous sections. From this, it can be concluded that there are four categories of blockchain implementations.

3.2.1.1.- Public and permissionless blockchains

In this category, the blockchain members send transactions using pseudonyms. For instance, in Ethereum users are identified by Ethereum addresses, which are the result applying several times a cryptographic functions to their private key. There are no pre-requisites for a node to be part of the network. In some deployments, resources are required to be able to propose an append action to the blockchain. In practice, these kinds of blockchains require a *proof-of-work* consensus protocol, or similar. There is no trust between the participant nodes, and neither does a penalty exist for nodes in case of them attempting fraud. Scalability here is undermined, due to the inherent trust model, but the solutions offer quite substantial integrity of information.

This category is not adequate for the SOTER proposition. SOTER platform must enforce identification and authentication of the blockchain nodes. In the case of a node acts maliciously and commits fraud, network capabilities should offer a reliable mechanism to identify these actions and conduct forensics tasks to collect evidences. This is especially relevant in the extended financial sector, where fraudulent action may have severe consequences.

Public and permissionless networks have specific characteristics, such as pseudonymity, lack of a central organization, censorship resistance and multiple, public copies of the ledger, which contribute to both availability and redundancy, but these characteristics are not wholly desirable for a project such as SOTER.

3.2.1.2.- Public and permissioned blockchains

This category integrates the public access characteristic of Public Blockchains, as a new kind of “Public Good” but stablishing the set of minimums obligations and restrictions to the participants at the time they deploy, write and read information of the network. Those

obligations and restrictions should be general and not restrictive, complying with anti-trust principles.

The permissioned component implies that only a close list of identified nodes is able to process transactions and add new blocks. Those nodes should be part of a legal entity (Association, Cooperative, LLC, or other limited liabilities structure) with legal responsibility to respond to liabilities, bad functioning or improvements. In this category, there is a restricted set of nodes which interact with the network and this information is disclosed. If individuals or organizations want to join a node to the blockchain, they provide credentials which enable membership. These credentials are approved by a centralized actor or by the community, and they can have different levels of assurance about the identity of the credential requestor.

Validators do not compete for adding a block to the chain, as happens in permissionless public blockchains. The motivation for them in the addition of blocks to the chain can reside in the fact that some nodes possess a greater amount of digital assets represented in the blockchain than others. By this reason, PoS consensus is used. It can be also possible that all nodes have the same interests in the proper functioning of the network. In these cases, the BFA consensus mechanism fits better. If members are clearly identified, they can be sanctioned or excluded in case of fraud. A contractual protection is required for this reason.

To guarantee stability, resiliency and decentralization of these networks, the institutions in charge of permissioned nodes have to comply with certain obligations (warranties, insurances, legal agreements, etc.) and should have to prove a minimum profile in terms of technical capability and infrastructure.

Examples of these type of networks are European Blockchain Services Infrastructure (EBSI), ALASTRIA, and LACChain (Lead by the Interamerican Development Bank).

A solution for a project such as SOTER, which is involved in the extended finance sector, requires a permissioned blockchain where it is possible to grant different kinds of access to the actors of the platform. It could be possible that auditor and regulatory bodies need to be represented by a node that has full read access to the information, in order to verify aspects of legality and compliance. Other actors, such as end-users may interact with the platform through a limited access node. These different access level can be achieved using specific credentials in a permissioned blockchain.

It should be noted that a public blockchain, which contains personal data, does not offer one of the main characteristics that a GDPR-compliant solution should have: information confidentiality for customers. Although the blockchain does not contain personal data, online identifiers[24] may be used to profile users and subsequently used to link them through off-chain identifying data. Careful consideration must be given to these sorts of deployments, as analysis must be completed on how feasible it may be to create an identifying link between data (hashes, addresses, tokens, transactions, etc) and the data subject.

3.2.1.3.- Private and permissioned blockchains

The nodes involved in this sort of blockchain must obtain a license and/or verified credential to operate as part of the network. The information is kept private, and access is only granted to those who have received access rights. These two characteristics are desirable for the SOTER project, in terms of confidentiality and control access. There is mutual trust between the nodes, which are operated by different organizations, so identification must be established in order to ensure there is liability and accountability with respect to information stored on the blockchain.

3.2.1.4.- Private and permissionless blockchains

In permissionless networks an authoritative party does not exist, so any node can join the network and send transactions[25]. Data privacy is achieved with smart contracts, which define who is allowed to invoke smart contract methods and obtain access to the data. Every time that a smart contract is instantiated, a private chain is automatically created associated with it. Read permission can be granted to organizations or specific persons.

[24] Recital 30 EU General Data Protection Regulation (EU-GDPR). Privacy/Privazy according to plan. (n.d.). Retrieved January 28, 2020, from https://www.privacy-regulation.eu/en/recital-30-GDPR.htm

[25] Daniels, A. (n.d.). The rise of private permissionless blockchains. Retrieved January 27, 2020, from https://medium.com/ltonetwork/the-rise-of-private-permissionless-blockchains-part-1-4c39bea2e2be

These networks have a few and limited use cases. For instance, LTO Network [26] is '*purposely build for verifying information and to support Live Contracts and the private event chains*'. Considering SOTER project requirements, private and permissionless blockchains are not suitable because they do not offer consensus and data sharing between consortium members.

3.2.2.- Consensus protocol.

Bearing in mind that the architecture for the SOTER platform should be a private and permissioned blockchain, protocols that are required for public and permissionless blockchains should be discarded. *Proof-of-Work* requires a high computation efforts and performance in the nodes, and should be demanded only by public blockchains. *Proof-of-Stake* and *Proof-of-Importance* do not offer enough neutrality and could allow some nodes to have large influence on the network. Protocols based on Byzantine Agreements should be used here. The SOTER project also requires an implementation which provides the possibility of balanced capacity rotation, so protocols as Raft and Istanbul Byzantine Fault Tolerance are suitable. If several organizations are present in the blockchain, it is required that a consensus agreement should be reached equally and fairly.

3.2.3.- Identification and authentication processes.

As stated in previous sections, a permissioned blockchain offers characteristics that a project involved in the financial industry requires. In this kind of blockchain implementation, a control layer is provided to govern the actions that allowed participants can perform. This layer enforces the policies managed by the blockchain governance body. Participants will operate in a trusted environment where their identities are known. To make it possible, it is necessary to establish an identification service that allows entities to authenticate themselves.

End-users will also need a user-centric, secure and legally compliant decentralized digital identity system. The digital identity of a user will inform an entity about his or her *identity attributes,* which are discrete pieces of information linked to a specific user. It should be built upon a decentralized identity paradigm, which can be achieved shifting most of the

[26] LTO.network. (n.d.). Lto network whitepaper. Retrieved from https://ltonetwork.com/documents/LTO Network - Technical Paper.pdf

capabilities to a user's hands, or at least trusting in decentralized methods and cryptographic algorithms. This identity system should provide end-users a digital sovereign identity, where the user will be the absolute owner of his or her personal data. He or she will manage the access to the information along with the possibility of sharing it. In *The path to Self-Sovereign Identity* [27], Allen defined ten principles of Self-Sovereign Identity:

- **Existence** - People have an independent existence.
- **Control** - People must control their identities.
- **Access** - People must have access to their own data.
- **Transparency** - Systems and algorithms must be open and transparent.
- **Persistence** - Identities must be long-lived.
- **Portability** - Information and services about identity must be transportable.
- **Interoperability** - Identities should be as widely usable as possible.
- **Consent** - People must freely agree to how their identity.
- **Minimization** - Disclosure of claims must be minimized.
- **Protection** - The rights of individual people must be protected.

All these properties suit the SOTER project requirements, so it is reasonable to explore the concept of Self-Sovereign Identity (SSI) and apply its principles to the identity framework that will govern the platform.

The World Wide Web Consortium (W3C) is an international community whose mission is to lead the WWW to its full potential, by publishing protocols, standards and guidelines to assure the proper growth of the Web. The W3C has created two different Working Groups:

- Verifiable Credentials (VC) Data Model

[27] Allen, C. (n.d.). The Path to Self-Sovereign Identity. Retrieved January 12, 2020, from http://www.lifewithalacrity.com/2016/04/the-path-to-self-soverereign-identity.html

- Decentralized Identifiers (DIDs)

They have developed specifications[28] [29] which are under active development.

It is relevant to mention here that in the interest of the European Blockchain Partnership (EBP) and the standardization of a citizen digital identity framework employing a SSI model. 21 Member States and Norway signed in 2018 a Declaration to create the EBP, and they are cooperating in the establishment of a European Blockchain Services Infrastructure (EBSI)[30]. The implementation of EBSI for an initial set of cross-border digital public services started in 2019 and it will have a minimum viable product in February 2020. These services include audit documents, cross-border certification of diplomas and a European Self-sovereign Identity initiative. This identity service will be built on eIDAS and it will offer a framework to provide cross-cutting capabilities. Therefore, the SSI will be an identity model and a reference for the SOTER project and the SOTER committee board should carefully monitor EBSI reports on the SSI framework. EBSI identity model follows the VC Data Model and DID specifications. A summary of the guidelines developed by these working groups is provided below.

VC Data Model specification provides a framework to express credentials on the Web in a "*cryptographically secure, privacy respecting and machine-verifiable*". A credential can represent all the information that a physical credential represents. It might be information related to identifying its subject (a name or an identification number), to the issuing authority (a government or a certification body), to information about the type of credential (a driving license) or to specific attributes being asserted by the issuing authority about the subject (the classes of vehicle entitled to drive). Some cryptographic techniques, such as digital signatures, are added to credentials to make them verifiable which ensure the credential is tamper-evident and trustworthy.

[28] W3C. (2019). Verifiable Credentials Data Model 1.0. Retrieved January 12, 2020, from https://www.w3.org/TR/vc-data-model/

[29] W3C. (n.d.). Decentralized Identifiers (DIDs) v1.0 First Public Working Draft. Retrieved January 12, 2020, from https://www.w3.org/TR/did-core/

[30] European Blockchain Services Infrastructure: https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/ebsi

The specification introduces some actors, which are presented below (see Figure 2), described as:

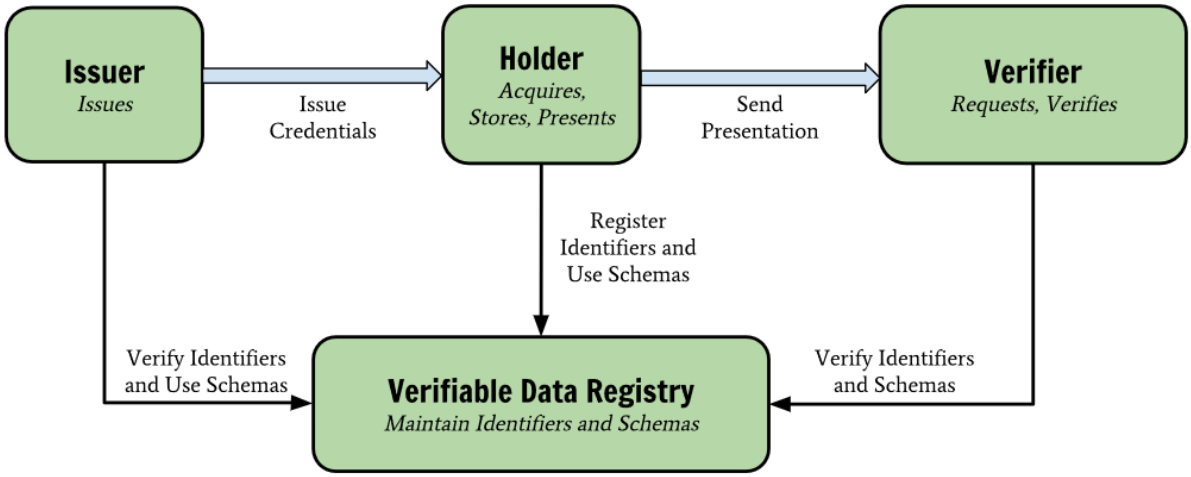

Figure 2. Verified Credentials Framework

The roles and information flows forming the basis for VC specifications:

- **Holder:** A entity that possesses one or more verifiable credentials and can generate verifiable presentations from them.
- **Issuer:** A entity that asserts information (claims) about subjects, and creates verifiable credentials from these claims and sends them to a holder.
- **Subject:** an entity (end-user or organization) about which claims are made. The role of the holder usually is the same than the subject, but it in some cases a holder can keep the credential of a subject (for example, a parent and a child).
- **Verifier:** An entity who receives, verifies and processes credentials.

For the purpose of understanding the concept of a verifiable credential, an example is detailed hereunder.

As it can be seen, a credential contains information about a subject which is digitally signed by an issuer:


```
{
 "@context": [
  "https://www.w3.org/2018/credentials/v1",
  "https://www.w3.org/2018/credentials/examples/v1"
 ],
 "id": "http://example.gov/credentials/3732",
 "type": ["VerifiableCredential", "UniversityDegreeCredential"],
 "issuer": "https://example.edu",
 "issuanceDate": "2010-01-01T19:73:24Z",
 "credentialSubject": {
  "id": "did:example:ebfeb1f712ebc6f1c276e12ec21",
  "degree": {
   "type": "BachelorDegree",
   "name": "Bachelor of Science and Arts"
  }
 },
 "proof": {
  "type": "RsaSignature2018",
  "created": "2018-06-18T21:19:10Z",
  "proofPurpose": "assertionMethod",
  "verificationMethod": "https://example.com/jdoe/keys/1",
  "jws": "eyJhbGciOiJQUzI1NiIsImI2NCI6ZmFsc2UsImNyaXQiOlsiYjY0Il19
   ..DJBMvvFAIC00nSGB6Tn0XKbbF9XrsaJZREWvR2aONYTQQxnyXirtXnlewJMB
   Bn2h9hfcGZrvnC1b6PgWmukzFJ1IiH1dWgnDIS81BH-IxXnPkbuYDeySorc4
   QU9MJxdVkY5EL4HYbcIfwKj6X4LBQ2_ZHZIu1jdqLcRZqHcsDF5KKylKc1TH
   n5VRWy5WhYg_gBnyWny8E6Qkrze53MR7OuAmmNJ1m1nN8SxDrG6a08L78J0-
   Fbas5OjAQz3c17GY8mVuDPOBIOVjMEghBlgl3nOi1ysxbRGhHLEK4s0KKbeR
   ogZdgt1DkQxDFxxn41QWDw_mmMCjs9qxg0zcZzqEJw"
 }
}
```


The entities involved in verifiable credentials are represented and identified by the framework provided by the Decentralized Identifiers specification. It takes advantages of blockchain technologies to provide a fully decentralized identity data model. Here, entities are identified by decentralized identifiers (DIDs). A DID is a pointer to a DID Document, which

contains one or more service endpoints for interacting with the entity identified by the DID. Following Privacy by Design principles, one entity can create as many DIDs as it needs to establish a secure separation of contexts.

DID methods are the different mechanism by which a DID and its related DID Document are created, read, updated or deactivated in a specific blockchain. Each DID method should be defined using DID method specifications.

Thanks to the identity management provided by DIDs, the dependence on centralized registries or hierarchical PKI is eliminated.

As was completed before, an example of a DID and DID documentation are provided for a better understanding. Hereunder an example of a DID is shown. It consists of a URL scheme identifier (did), an identifier for the DID method (example), and a DID method-specific identifier (123456789abcdefghi):

- did:example:123456789abcdefghi

The following structure is an example of a DID Document provided by the specification. It contains the DID relative to this DID Document. It also includes a service endpoint to interact with the DID subject, a public key which is used by the subject to interact with others. This public key can be used for digital signatures, encryption and other cryptographic operations, which are the fundamentals for authentication or to establish a secure connection with other services. Using these keys, a DID subject can cryptographically prove that they are associated with a DID.

```
{
 "@context": "https://www.w3.org/ns/did/v1",
 "id": "did:example:123456789abcdefghi",
 "authentication": [{
  // used to authenticate as did:...fghi
  "id": "did:example:123456789abcdefghi#keys-1",
  "type": "RsaVerificationKey2018",
  "controller": "did:example:123456789abcdefghi",
  "publicKeyPem": "-----BEGIN PUBLIC KEY...END PUBLIC KEY-----\r\n"
 }],
```

```
 "service": [{
  // used to retrieve Verifiable Credentials associated with the DID
  "id":"did:example:123456789abcdefghi#vcs",
  "type": "VerifiableCredentialService",
  "serviceEndpoint": "https://example.com/vc/"
 }]
}
```

As it is stated in DID Data Model specification[31], this framework, using the advantages of blockchain technologies, provide the following benefits:

Decentralization: Eliminate the requirement for centralized authorities or single point failure in identifier management, including the registration of globally unique identifiers, public verification keys, service endpoints, and other metadata.

Control: Give entities, both human and non-human, the power to directly control their digital identifiers without the need to rely on external authorities.

Privacy: Enable entities to control the privacy of their information, including minimal, selective, and progressive disclosure of attributes or other data.

Security: Enable sufficient security for relying parties to depend on DID documents for their required level of assurance.

Proof-based: Enable DID subjects to provide cryptographic proof when interacting with other entities.

Discoverability: Make it possible for entities to discover DIDs for other entities to learn more about or interact with those entities.

Interoperability: Use interoperable standards so DID infrastructure can make use of existing tools and software libraries designed for interoperability.

[31] W3C. (n.d.). Decentralized Identifiers (DIDs) v1.0 First Public Working Draft. Retrieved January 12, 2020, from https://www.w3.org/TR/did-core/

> **Portability:** Be system and network-independent and enable entities to use their digital identifiers with any system that supports DIDs and DID methods.
>
> **Simplicity:** Favour a reduced set of simple features to make the technology easier to understand, implement, and deploy.
>
> **Extensibility:** Where possible, enable extensibility provided it does not greatly hinder interoperability, portability, or simplicity.

For all these desirable characteristics, SSI data model represents the guidelines which the identity framework provided in SOTER project should follow. Along with the development of the project, SOTER members will analyse and discuss:

- How it is possible to link the DID schematic with the identity schematic provided by eIDAS
- How the interaction of the identity data model with the on-boarding platform will be eIDAS compliant
- How the SSI model will be compliant with GDPR

The outcomes of these analysis will be presented in the second part of this deliverable in M19.

3.3.- Data Layer

As stated in the previous chapter, the risks in this layer are associated with the confrontation between blockchain characteristics such as transparency and shared data access, and the desirable levels of data protection. The information stored on the blockchain should be protected by privacy and confidentiality, while also being GDPR compliant. Some recommendations will be provided in this section that should be taken into consideration during the design of the platform.

First, we should establish a definition of what *personal data* is with respect to GDPR. We can find the definition within Article 4:

> 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.

This definition clearly makes a reference to data that can identify a natural person (full name, national id number, telephone number or complete address), but it also refers to information, that combined with other pieces, could potentially identify a person (city, country, age or gender).

GDPR also states, within Recital 30:

> Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags. This may leave traces which, in particular when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them.

This recital includes a wider definition of *personal data* than the one provided by Article 4, including information such as device identifiers, cookies and/or IP addresses. Taking into account this recital is it necessary to carefully review what kind of information will be stored in the blockchain during the on-boarding design processes.

If there is a requirement to store personal data or private information in the blockchain, there are a set of techniques or mechanisms that may be applied:

- **Hashing:** Generate a fixed-length value from a string of text using a mathematical function. If the algorithm is secure, it is computationally hard to reverse a hash function to find a different input with the same hash. It is important that the length of the generated hash is long enough (at least, longer than the length of the possible entries to the hash function) to avoid collisions. If the set of possible entries to the hash function is already known, an attacker could guess the content of the information that had been hashed. This could happen it the attacker knows the set of possible value, or the pattern used to form a string.

The Spanish Data Protection Board has developed some guidelines[32] for using hashes in a secure manner.

- **Encryption:** Using a secret key to cipher information is a useful technique to obtain privacy and confidentiality, but there are some risks to consider. If the key is shared between blockchain members, it could result in a leak of information. Besides that, cryptanalysis techniques and the advent of quantum computation pose threats to the safety and security of current algorithms, which could make public the encrypted information.

- **Anonymisation:** The process of removing personal identifiers, both direct and indirect, that might lead to identify an individual.

- **Pseudonymisation:** the processing of personal information in a way that it can no more be connected to a subject without the use of additional data. This additional data is kept separate to ensure that the processed data cannot be attributed to an individual. This action is reversible, which means that the processed and seemingly anonymous data can be linked to an individual if an attacker can obtain additional data about the pseudonymisation process.

- **Off-chain storage:** This procedure collects personal data in an off-chain and secure storage. A link to this off-chain information is stored in the blockchain, and the subject of the personal data holds a private key to grant access to the off-chain data.

The design of SOTER project will consider all of these procedures and techniques, along with the reports and guidelines from expert committees such as *The Article 29 Working Party*, which published a 2018 report about processing of personal data[33]. It is highly recommended the "Solutions for a responsible use of the blockchain in the context of personal data",

[32] EDPS. (2019). Introduction to the hash function as a personal data pseudonymisation tecnique. Retrieved from https://edps.europa.eu/sites/edp/files/publication/19-10-30_aepd-edps_paper_hash_final_en.pdf

[33] Article 29 Working Party. (2014). Opinion 05/2014 on Anonymisation Techniques. Working Party Opinions, (April), 1–37. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

published by the French Data Protection Board[34]. As one of the strengths of SOTER project is the establishment of a user data sovereignty over a secure framework, off-chain storage (user controlled) is an appealing mechanism to be chosen for the project. This procedure lacks the potential weaknesses that other mechanisms might have, and it is user-centric, in terms of the data subject can grant or deny the access to the personal information as and when required.

There is also one feature that should be implemented in the data layer to complement the blockchain functionality. Most of the available blockchain implementations offer audit and logging capabilities processes to store and review write accesses, because they are performed through transactions. Therefore, this log information can be accessible at any time and integrity is preserved. Nevertheless, read accesses are not usually stored in blockchain beyond logs written in files, where integrity is not guaranteed. From a legal perspective, read accesses are at least as valuable as write accesses.

3.4.- Application Layer

In the previous chapter some security guidelines were addressed to be followed in the development of the blockchain smart contracts. As any piece of software, they should include an auditing process, penetration tests and vulnerabilities scans. There are a number of related tasks within the SOTER DoA that related specifically to these requirements. They may be read within Section 1.4.

It will be highly recommended to follow and, if possible, to join security initiatives such as the Open Web Application Security Project (OWASP). This organization is developing a Blockchain Security Framework[35] which aids understanding of the requirements needed to ensure maximum security at each stage of the product development, focusing on code and smart contracts.

[34] CNIL. (2018). Blockchain. Solutions for a responsible use of the blockchain in the context of personal data. CNIL Report, 10. Retrieved from https://www.cnil.fr/sites/default/files/atoms/files/blockchain.pdf

[35] OWASP. (2019). Blockchain Security Framework - OWASP. Retrieved from OWASP website: https://www.owasp.org/index.php/Blockchain_Security_Framework

3.5.- Infrastructure Layer

At this level, it is mandatory to set a secure environment for the cryptographic keys related to the blockchain network. If the nodes of the network are identified by a pair of private and public keys, security can be enforced using hardware security modules (HSMs). They are physical devices that safeguard and manage private keys, and they can ask for a passphrase and/or a physical token such as a cryptographic card to use the keys. HSMs provide a controlled and secure environment to execute cryptographic functions, such as keys generation or digital signatures. A HSM must be used explicitly to guard these keys at every phase of their life cycle. A secure channel is stablished between the application (in this case, the blockchain node) and the HSM where the private key is stored. Access to these keys must be audited and there must be a closed list of users who can interact with the HSM with administration proposes. US National Institute of Standards and Technology (NIST) provides guidelines and recommendations for key management.[36] NIST also maintains a list of approval status of algorithms and key lengths for cryptographic proposes.[37]

The blockchain network that support SOTER project should be audited on an iterative basis to detect security weaknesses. It is recommended to develop an audit program which periodically performs vulnerabilities scans, to detect weaknesses in the blockchain network applications and protocols that may be exploited. The life cycle of the found vulnerabilities must be managed using a tracking application in order to point a responsible and a deadline to solve them.

It is highly recommended to perform a security analysis of the blockchain network to reduce the attack surface in case any node is compromised. We should take into account that the blockchain network interacts with external applications. The network should follow a “Privacy by Design” and “least privilege” principles to minimize the exposed endpoints. This can be achieved with the proper use of firewalls and with the hardening of the blockchain nodes, assuring that only the necessary ports and services are enabled. This analysis must include also a network segmentation plan, to stablish a division of a wide network into security zones.

---

[36] Barker, E. (2016). Recommendation for Key Management. NIST Special Publication 800-57, 1–142. https://doi.org/10.6028/NIST.SP.800-57pt3r1

[37] Barker, E., Roginsky, A., Locke, G., & Gallagher, P. (2011). Recommendation for Transitioning the Use of Cryptographic Algorithms and Key Lengths. NIST Special Publication, (January), 800–131. Retrieved from https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-131Ar2.pdf

For instance, one of the most well-known blockchain, Hyperledger Fabric, is composed of different type of nodes[38] which perform activities that should be executed in separated networks.

4.- Blockchain implementation chosen for SOTER Project

As it is stated in the SOTER DoA, the onboarding platform will rely on the ALASTRIA network. This is a national Spanish blockchain network, where multiple services providers from a number of sectors are present, including utilities, banking, and the public sector. The reason for this is to take advantage of the developments and evolution gained within the ALASTRIA network, in order to further the blockchain technology evolution and reduce the risks that can arise regarding the use of this disruptive technology.

In this section, the blockchain implementation of the ALASTRIA Network will be described and analysed.

4.1.- Description

The ALASTRIA network is a multisectoral consortium based in Spain. It has adopted the legal form of a non-profit association, and their main objective is to promote the digital economy through the establishment of decentralised ledger technologies. It is open to all types of companies and organizations, in order to reach all sectors and contribute to the creation of a diverse as possible ecosystem. It was launched in May 2017, and since that, it has been growing and focusing on the creation of basic services regulated and adapted to Spanish legislation.

ALASTRIA states that they built their platform over three pillars:

[38] https://hyperledger-fabric.readthedocs.io/en/release-1.4/arch-deep-dive.html#system-architecture

- **The Association**: a "non-profit association that promotes the digital economy through the development of decentralised ledger technologies/Blockchain" [39]
- **The Network**: Built over Quorum (Ethereum-based architecture) a public permissioned network.
- **ALASTRIA_ID Project**: a digital identity model proposed by the Association.

In the scope of this document, a research about the network and ALASTRIA_ID project will be done in the next sections.

4.2.- The Network

The ALASTRIA current network is built on Quorum, which is an Ethereum-based blockchain architecture. As it is stated on their website presentation, ALASTRIA is "*an agnostic blockchain platform (we do not trust our development to a single platform), so work has started to create two types of new networks, one based on Parity and one based on HyperLedger Fabric*"[40].

4.2.1.- Quorum

This security analysis will be focused on the operational ALASTRIA network, entitled Telsius.

Finance is considered one of the first industries that will be disrupted by blockchain technology. With this aim, Quorum was developed by J.P. Morgan, and focused on the financial industry. It is built upon the base code of the Ethereum blockchain. This means that software elements, called Smart Contracts, can be instantiated and executed on a replicated and shared ledger. As a result, business logic is distributed across nodes, improving efficiency and lowering costs comparing to traditional business enterprise systems. The fact that Quorum is based on the Ethereum codebase is beneficial for the SOTER project. First, Ethereum is an open source project, which provides transparency, and allows for easier identification of problems and/or performance issues due to the code being open and public.

[39] https://alastria.io/en/
[40] https://alastria.io/en/la-red/

Ethereum has been implemented in many projects and is considered one of the most mature blockchain architectures. The mainnet has been operational since July 2015, and, since that, a myriad of unit tests and security programs have been evaluated on it. Also, the Ethereum community manages a bounty program, which rewards developers for finding security vulnerabilities in the codebase. The community is formed by an ecosystem of developers, tools and applications.

However, the Ethereum network lacks data privacy, and the data managed by the smart contracts is exposed to the public. Based on this, Quorum provides a different techniques to improve security and data privacy on their network. These characteristics, which are requirements for the SOTER project, will be explained below.

4.2.1.1.- Consensus protocols available in Quorum

The Quorum foundation is concerned with data privacy. It takes advantage of cryptographic techniques to prevent unauthorised entities from viewing sensitive data, except for the actors involved in a transaction. To reach this objective, Quorum is based on a single blockchain and combination of smart contracts and modifications to the original Ethereum codebase. Smart contracts need to achieve certain privacy levels, as they are pieces of software that build on segmentation of private data within Quorum. Ethereum is therefore modified in terms of the block validation process: public transactions are validated by all nodes, but private transactions are skipped by the nodes that do not take part in the transaction.

Considering this, Quorum's state database is segmented: there is a public and a private database. All nodes are in perfect consensus with the public database, while private databases differ: each node only stores the private database that it has permission to access. Although a node does not store the state of the whole database (the node only keep the private data that it has access to) the actual distributed blockchain and all the transactions are fully replicated in all nodes. In the next figure, taken from Quorum Whitepaper[41], the concept is explained:

[41] quorum-docs/Quorum Whitepaper v0.2.pdf at master · jpmorganchase/quorum-docs · GitHub. (n.d.). Retrieved November 9, 2019, from https://github.com/jpmorganchase/quorum-docs/blob/master/Quorum Whitepaper v0.2.pdf

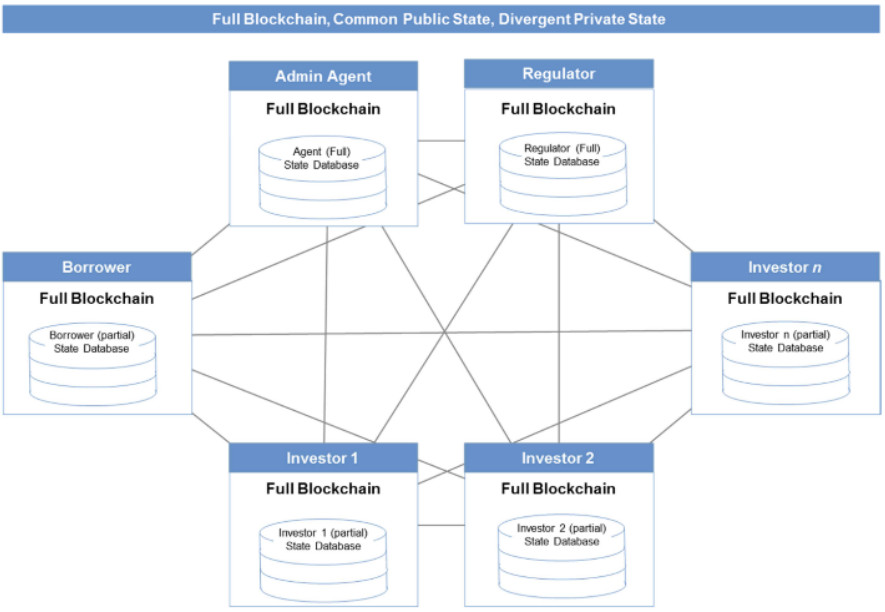

Figure 3. Quorum Network

Data privacy is one of the key topics of the SOTER Project. It has a high importance for the project because citizens have to trust digital transactions. Also, the SOTER Project has to be fully compliant with EU General Data Protection Regulation. For this reason, the privacy offered by Quorum blockchain fits with the SOTER requirements.

4.2.1.2.- Data Privacy in Quorum

Focusing on consensus protocols, Quorum implements two mechanisms[42]. The protocol that is available by default is Raft, but Istanbul BFT can also be selected.

First, it has to be decided whether Byzantine fault tolerance is a requirement or not. This means that the system has the ability to keep functioning even when some nodes are not in agreement (consensus) with the majority.

[42] quorum-docs/Quorum Whitepaper v0.2.pdf at master · jpmorganchase/quorum-docs · GitHub. (n.d.). Retrieved November 9, 2019, from https://github.com/jpmorganchase/quorum-docs/blob/master/Quorum Whitepaper v0.2.pdf

As it has been explained in previous sections, Raft offers a leader/follower model. There is a single leader for the entire group of nodes, which is the only one that should generate new blocks. Raft is not a Byzantine fault tolerant algorithm: the nodes trust the elected leader.

IBFT is a Byzantine fault tolerant solution. As was explained previously, IBFT uses a group of validator nodes to ensure integrity of each block. Two thirds of them are required to sign the block before inserting it to the chain. This ensures it is difficult for nodes to commit fraud.

In terms of performance, IBFT offers immediate transaction finality, due to the fact that there is only one block proposed at a given time. The effort needed to construct and validate the block is reduced, compared to other mechanisms, which has the effect of increasing the throughput (measured in transactions per second) of the network.

It is also necessary to note that IBFT provides a rotational mechanism with regards the leadership of the group, ensuring crash tolerance and also ensuring no one node has disproportionate influence over the blockchain.

In view of above, IBFT is the most suitable consensus mechanism that should be implemented, in terms of byzantine fault tolerance, crash tolerance, performance and leadership rotation, bearing in mind the requirements of the SOTER project.

In the whitepaper of Quorum, it is stated that tests have demonstrated a throughput of "dozens to hundreds of transactions per second", so in terms of performance it could be sufficient to develop a minimum valuable product, but not for a production environment platform. Research is required to find out if this throughput could be leveraged by increasing the number of nodes. Apart from this, the Quorum network offers enhanced security capabilities. However, a source from ALASTRIA Blockchain Ecosystem confirms:

> "it should also be noted that after the initial push of commits from the open source community, in recent years, there has been less community activity and these types of issues have been left in the backlog for some time. There is heavy competition in the Enterprise Ethereum area and platforms such as BlockApps (similar to Quorum, but with a management and smart contract layer called Strato that simplifies integration and

development greatly) and Pantheon (now Hyperledger Besu — Officially an Ethereum Client) share this market space."[43]

In addition, ALASTRIA is evaluating the deployment of another network in Hyperledger Besu. This network will be discussed in the next section.

4.2.2.- Hyperledger Besu

Hyperledger Besu [44] is an Ethereum client designed for both public and private permissioned networks. It is an evolution of Pantheon, an open source solution provided and developed by ConsenSys. Besu is considerate as an improved for public-permissioned networks in the Ethereum open source space, also because the interoperability with Hyperledger Fabric.

A high-level architecture of Besu can be shown below (See Figure 4).

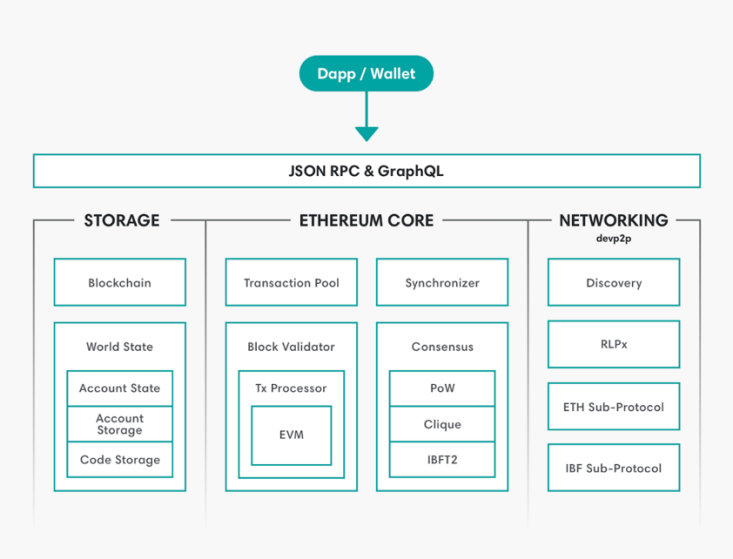

Figure 4. Hyperledger Besu Architecture [45]

[43] Creer, D. (n.d.). Comparison of DLT platforms - Alastria Blockchain Ecosystem. Retrieved from https://medium.com/@alastria_es/comparison-of-dlt-platforms-be84950d339d

[44] https://www.hyperledger.org/projects/besu

[45] https://besu.hyperledger.org/en/stable/Concepts/ArchitectureOverview/

4.2.2.1.- Consensus protocols available in Besu

Besu[46] includes a wide range of consensus algorithms, such as PoW, IBFT, IBFT 2.0 and Clique. IBFT 2.0[47] is an improved version of its predecessor in terms of key safety, extended fault tolerance and offers a higher performance for enterprise networks. Here, transactions and blocks are validated by authorized nodes, known as validators, which take turns to create a block. If more than two thirds of the validators sign the block (a super majority of them), it can be inserted into the chain. This is a modification in IBFT 2.0, which addresses the issue of Byzantine nodes being able to reach agreement in IBFT 1.0[48]. If the proposer is acting maliciously, then it is possible that it can have different blocks committed at the same height of the blockchain by reaching consensus with two different sets of validators. For instance, as we can see in Figure 5, if there are five validators, IBFT 1.0 requires agreement of only 3 nodes for the next block to be added to the blockchain. In this case, a malicious proposer can reach consensus on different blocks with two distinct sets of validators. In IBFT 2.0 it would be necessary to reach consensus with four validators agreeing.

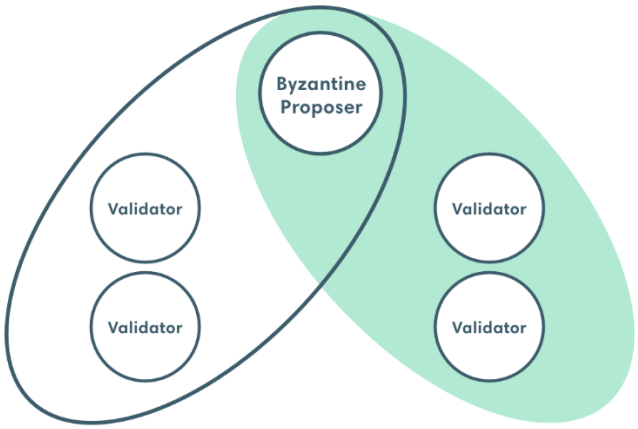

Figure 5. IBFT 1.0 issue of byzantine nodes [48]

[46] https://besu.hyperledger.org/en/stable/Concepts/Consensus-Protocols/Overview-Consensus/
[47] https://besu.hyperledger.org/en/stable/HowTo/Configure/Consensus-Protocols/IBFT/
[48] Rubino, G. (n.d.). Another day, another consensus algorithm. Why IBFT 2.0? Retrieved from https://pegasys.tech/another-day-another-consensus-algorithm-why-ibft-2-0/

There is also a mechanism to add or remove validators, which need a majority vote (more than a half of the validators). It requires at least 4 validators to be byzantine fault tolerant. In case of a failure, it requires at least two thirds of validators to be operating to create blocks. For instance, in a network of six nodes, two unresponsive nodes are tolerated. If they are less than six nodes, only one unresponsive node is tolerated.

In the Clique mechanism protocol[49], nodes that can add a block are called sealers. When a sealer signs a block, it is not allowed to seal a next fixed number of blocks. If there are N sealers and they can sign 1 block out of K, at any point in time there are [(N – K)+1] sealers allowed to sign. To avoid racing for blocks, every sealer waits a random time to release a new block. It ensures that forks are rare. A fork in a blockchain occurs when two nodes sign a block and there is a bifurcation in the chain. Nodes should be aware of forks and add wait a random time before signing. By this reason, IBFT 2.0 is preferred over Clique for SOTER Project. IBFT 2.0 has immediate finality because there are no forks and all valid blocks are automatically appended to the blockchain[50]. Also, Besu documentation suggests that, for systems that require data privacy, they "*recommended using a network with a consensus mechanism supporting transaction finality. For example, IBFT 2.0*"[51].

4.2.2.2.- Data Privacy in Besu

Hyperledger Besu builds privacy upon a private transaction manager such as Orion[52]. It is an application that create and maintains cryptographic key pairs, stores privacy group details, and provides an API for communication. Each Besu node needs an Orion node, which encrypts the information and distributes it in point-to-point communications to the destination Orion node. An example is provided in below (see Figure 6).

[49] https://github.com/ethereum/EIPs/issues/225
[50] https://besu.hyperledger.org/en/stable/Concepts/Consensus-Protocols/Comparing-PoA/
[51] https://besu.hyperledger.org/en/stable/Concepts/Privacy/Privacy-Overview/
[52] http://docs.orion.pegasys.tech/en/stable/

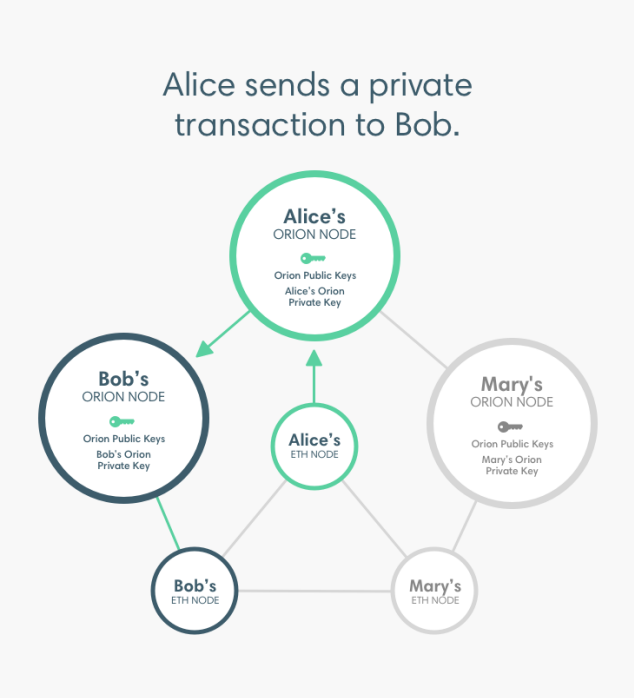

Figure 6. Private communication in Hyperledger Besu

Besu and Orion nodes both have a pair of cryptographic keys. Besu nodes' private keys are used to sign a transaction. Then, the transaction is encrypted with the destination Orion node public key. The mapping between keys and destination addresses should be stored off-chain.

Besu enables the creation of privacy groups. These are groups of nodes among whom private encrypted information is shared. To permit this, Besu maintains a public database that is shared between all nodes, and private databases, one for each privacy group.

4.2.3.- Recommendations for the SOTER Project – Blockchain Platform

The Hyperledger Besu network is preferred over Quorum for the SOTER Project. Both frameworks offer IBFT consensus protocol, which is the most suitable, and also provide encryption techniques to ensure aspects of data privacy. As was mentioned before, a permissioned network is desirable for SOTER Project, which can be achieved in Quorum and Besu. However, the Quorum community seems less active, as stated before. It is relevant to consider that the European Commission, in charge of European Blockchain Services Infrastructure (EBSI) are considering the Besu Network as the most complete solution for the

EBSI project. It is also important to note that Besu is written in Java, the mainstream business programming language.

4.3.- ALASTRIA_ID Project

In order to finish the review of the three pillars of ALASTRIA network, ALASTRIA_ID project will be discussed in this sub-section. ALASTRIA_ID is the digital identity sub-system of the network. The association considers Identity particularly important, so an Identity Commission was established to define and create this sub-system [53]. It has pointed three lines of work: model standardization, creation of a reference and the implementation of the model. ALASTRIA considers that identity model should be legally binding and is built upon three premises:

- **Security**: Information related with the identify must be secure stored, with protection, persistence and minimization principles.
- **Controllability**: The user must control who can access to the information related with his or her identity.
- **Portability**: The user must be able to use her or his identity information in multiple devices and the transference of this information between identity providers must be feasible.

In addition, the identity model must be legally binding and be compliant with eIDAS, GDPR and money-laundering regulation, such as 5AMLD[54].

ALASTRIA_ID is based on the Self Sovereign Identity that was presented in the previous section. End-users will have the control of the operations and transactions associated with their identities. A smartphone application has been developed to be used as a wallet to store the private information (private keys) of the identity.

The model is based on the Figure 2 proposed by Verified Credentials Working Group where:

- The identity holders are citizens, the identity owners.

[53] https://alastria.io/en/estructura-de-alastria/
[54] https://medium.com/@alastria_es/as%C3%AD-avanza-el-proyecto-alastria-id-c206aa649770

- The verifiers are service providers that offer services to the citizens and require credentials to them.
- The issuers are the entities that are able to sign credentials to the citizens.

An end-user can request for credentials for the issuers, keep them in the wallet, and show them to the service provider when she or he asks for a service. A credential can be an identity credential, analogue to the national-ID, but also anything that an issuer can affirm about us, for instance, a drivers licence or a university diploma.

The credentials have different level of assurances, as the eIDAS regulation states. Here, the level of assurance depends on the importance or liability of the issuer: you can affirm who you are (low level), but a university (medium level) or the government (high level) can also do it.

When an end-user asks for a service to a provider, the provider sends a presentation request. A presentation is a collection of credentials. The presentation request must have a reason for the treatment of the information. Then, the user receives the presentation request, and if she or he agrees with the data treatment (GDPR compliant), she or he builds a presentation structure containing the credentials required.

If any of these credentials are not stored in the wallet, they must be requested to the proper issuer. Finally, the presentation is sent to the service provider, that verify them, and if all them are valid, the service is started.

End-users, service providers and issuers are identified by a DID following the DID specification presented in the previous chapter. The DID Document is stored in the blockchain. ALASTRIA_ID keeps in the blockchain the hashes of the issued credentials and the hashes of the presentations that have sent to the providers. In this way, the personal data is pseudo-anonymized.

In fact, ALASTRIA_ID stores four kinds of hashes and their status:

- A hash of a structure formed by a credential and the citizen DID. This is stored by the citizen, and only he or she can modify its status, in case she or she decides to delete it (GDPR compliant).
- A hash of a structure formed by a credential and the issuer DID. This is stored by the issuer, and only this entity can modify its status, in case it decides to revoke it.
- A hash of a structure formed by a presentation and the citizen DID. This is stored by the citizen, and only he or she can modify its status, in case she or she decides to delete it (GDPR compliant). If the presentation is marked as deleted, the service provider that had requested it before must delete it from their databases.
- A hash of a structure formed by a presentation and the service provider DID. This is stored by the service provider, and only this entity can modify its status, in case it decides to revoke it.

4.3.2.- Recommendations for SOTER Project – ALASTRIA_ID

The identity model provided by ALASTRIA_ID, as described in their documentation[55] claims to be GDPR compliant and follows eIDAS regulation. A deep analysis should be done with regards to its data model to assert whether this is true. It is necessary to find out how ALASTRIA will manage natural person identities to follow eIDAS regulation. This analysis will be covered in the next iteration of this whitepaper (M19).

[55] https://github.com/alastria/alastria-identity/wiki

5.- Conclusion

This document provides the first submission of the set of deliverables associated with *T3.3 – Block Chain Security Focus*. The deliverable is entitled *D3.5 – Blockchain Security Focus whitepaper (I).* The document provides a high-level overview of blockchain technology, with specific focus on aspects concerning blockchain security. The document provides an overview through a layered methodology, outlining aspects of the business, governance, data access, and consensus layers found within blockchain technology. It also provides information pertaining to identification and authentication processes, GDPR considerations, and details specific considerations regarding the application and infrastructure layers. Following on from this, the document outlines specific considerations of the SOTER project, outlining initial recommendations for the project. The document provides information regarding the proposed consensus mechanisms, platform architecture, integrated verified credential mechanisms, and comments on the specific platform considerations related to the ALASTRIA network, which is the proposed platform on which the SOTER Digital Onboarding Platform will be built.