A Quantifiable Trust Model for Blockchain-based Identity Management

Andreas Grüner, Alexander Mühle, Tatiana Gayvoronskaya, Christoph Meinel
Hasso Plattner Institute (HPI)
University of Potsdam, 14482, Potsdam, Germany
Email: {andreas.gruener, alexander.muehle, tatiana.gayvoronskaya, christoph.meinel}@hpi.uni-potsdam.de

***Abstract*—Removing the need for a trusted third party, blockchain technology revolutionizes the field of identity management. Service providers rely on digital identities to securely identify, authenticate and authorize users to their services. Traditionally, these digital identities are offered by a central identity provider belonging to a specific organisation. Trust in the digital identity mainly originates from the identity provider's reputation, organizational functioning and contractual obligations. Blockchain technology enables the creation of decentralized identity management without a central identity provider as trusted third party. Therefore, the derivation of trust in digital identities within this paradigm requires a distinct approach.**
In this paper we propose a novel general quantifiable trust model and a specific implementation variant for blockchain-based identity management. Applying the model, trust is deduced in a decentralized manner from attestations of claims and applied to the associated digital identity. This concept replaces trust with a central identity provider by aggregated trust into attestation issuers. Thus, promoting self-sovereign identities to be fit for purpose. The calculated numerical trust metric serves as independent basis for the definition of assurance levels to simplify and automate reasoning about trust by service providers without requiring a dedicated evaluation of a trusted third party.

***Index Terms*—Blockchain, distributed ledger technology, digital identity, self-sovereign identity, trust, identity management**

I. Introduction

The rise of blockchain technology on the grounds of Bitcoin's introduction [1] laid the foundation to eliminate the need for a trusted third party (TTP) in various domains [2]. Further steps in the development of blockchain technology drive its advancement from a straightforward peer-to-peer digital cash system to a general decentralized execution environment, which enables the implementation of new application concepts. In this regard, the domain of identity management is a dynamic research area for new approaches as it replaces TTPs with distributed consensus and data storage. Specifically, blockchain technology allows the implementation of decentralized identity management systems providing digital identities, that are not issued by a TTP. A decentralized digital identity that is under true control of the respective subject and that satisfies further criteria relating to security, controllability and portability is referred to as self-sovereign identity (SSI) [3] [4].

In traditional identity management, a digital identity and its attributes are issued by one or more central identity providers belonging to a specific organization. An organization in itself forms a trust domain. Service providers within the same organization naturally trust internal identity providers. By applying identity federation between organizational boundaries, the establishment of mutual trust is an inevitable prerequisite. The digital identity and its attributes derive trust directly from the identity provider and its organization [5]. Generally, due diligence of actions, contractual obligations and the reputation of the association affects the trust, that is attributed to the issued digital identity. An example is age verification. A statement about the age issued by a person themselves is in most cases insufficient and not relied on by the service provider. On the other hand, the personal identity card, which gives a date of birth and is state issued, is accepted. In this case the service provider attributes higher trust in terms of correctness to the identity card and the issuing state as compared to a statement given by the person. Altogether, based on trust in the identity and its characteristics, a service provider decides to offer services to the respective subject.

In contrast, an SSI is not issued by an identity provider as TTP and therefore, deriving trust from a central authority is not possible. However, service providers still have a strong demand for determining trust of an SSI and the corresponding attributes to offer services. To overcome this challenge, we propose a novel general quantifiable trust model and devise a specific implementation variant to calculate a trust metric independently from one central TTP that traditionally issues digital identities.

As the foundation we assume the attributes of an SSI are modelled as claims and attestations. A claim is a plain statement about a digital identity. An attestation is a statement from an entity to assert the correctness of a claim. Our novel trust model uniquely derives a trust value for a claim from its attestations and the respective attestation issuers. Furthermore, the trust values of the claims are aggregated to an overall trust value of the digital identity itself. The digital identity's entire metric and the claim-specific values enable easy and automated reasoning on trust decisions for service providers and facilitate the definition of qualitative assurance levels based on intervals. A dedicated trust evaluation of neither the identity provider nor the attestation issuer is required by service providers.

The rest of this paper is organized as follows. In Section 2 we provide an overview of the related work in this area and differentiating criteria. Background on trust in blockchain-based identity management is outlined in Section 3. In Section 4 we

describe in detail our proposed novel quantifiable trust model for blockchain-based identity management. Subsequently, in Section 5 we conduct a security analysis taking into consideration popular threat models against trust schemes. We discuss our devised trust model in Section 6, conclude the paper in Section 7 and present future work afterwards.

II. Related Work

Trust models mainly focus on two different areas. The first field is non-hierarchical communities (e.g. peer-to-peer networks or online feedback groups). The second field is identity management with assurance frameworks to specify trust in digital identities and its attributes. We analyse both areas and outline common and distinguishing factors compared to our quantifiable trust model.

In the first area of non-hierarchical communities peers build trust directly between each other. Therefore, these societies are structurally comparable to a network of self-sovereign identities. A classical application in the field of peer-to-peer networks is file sharing. Online feedback groups are also used to evaluate products and provide recommendations. The objective of the trust model is to reduce the number of corrupted files that are downloaded or the purchase of undesirable products. Trust building is based on prior experiences accumulated to make up reputation for a particular participant. EigenTrust is a leading reputation model in peer-to-peer networks [12]. In EigenTrust, peers compute local trust values based on transactions with neighbour peers. These transactions are classified into the categories successful and non-successful actions. Positively rated transactions increase the reputation of the respective peer, whereas negatively assessed transactions decrease the reputation. The local trust scores of each peer for every neighbour are accumulated to global values in order to provide a consolidated view of the network for all participants. In contrast, our model derives trust from attestations referencing claims of identities and is not based on previously gained experience. Claims and attestations are an essential part of an SSI.

Donato et al. [14] present particular web site ranking algorithms that are adapted to the peer-to-peer file sharing scenario in order to increase the efficiency of EigenTrust. Additionally, new attack models and a dishonest metric is proposed. Dishonesty describes negative reputation and is analogue to distrust. Distrust is not applied in our trust model to prevent the whitewashing attack. Appleseed [15] is a further trust propagation algorithm based on the concept of spreading activation models. Trust relationships are modelled in a graph-based network. Edges denote a trust energy flow from one node to another. In contrast to our trust model, the graph of Appleseed is not designed with regard to attestations, claims and digital identities. Thus, it does not reflect the self-sovereign identity model.

Overall, a significant difference between the described algorithms and our trust model is the utilization of experience-based reputation. In contrast, the foundation of our trust model is the aggregation and distribution of trust using attestation, claims and digital identities. Nonetheless, the described attack scenarios and requirements of the listed models are applicable to our trust scheme.

TrustMe [16] is a protocol to exchange trust values in a decentalized peer-to-peer network. The scheme does not focus on building trust values, however it targets the exchange of information in a decentralized way. In contrast, our trust model's objective is to derive trust values and it utilizes the blockchain network for data transfer. Therefore, TrustMe targets a different protocol layer.

Assurance frameworks are the second domain providing analogous approaches. Thomas et al. [5] [6] [17] propose a structure for attribute assurance to specify trust decisions on characteristics of a digital identity. A knowledge database, that is service provider specific, is utilized to store trust information about identity providers and issued attributes. This information declares whether the respective identity provider is trusted for a particular attribute. Ensuing, a binary trust decision is modelled as logical conclusion of expressions about the identity provider and the required attributes in a specific scenario. Compared to our trust model, the scheme of Thomas et al. differs with regard to the trust decision process and the generation of trust information. Our model offers a continuous trust score in a fixed interval as the basis for a decision instead of a binary logical conclusion. Additionally, our scheme aggregates trust information from different attestation providers to claims and, finally, to the digital identity compared with a distinct decision on identity provider and attribute basis. Furthermore, using the trust model of Thomas et al. a dedicated trust information database is needed on the side of the service provider. Overall, a greater flexibility and finer granularity is offered by our trust model including lower effort on evaluating trust information for the service provider.

An additional proposed trust scheme is the AttributeTrust [19] framework to determine confidence in attributes of various attribute providers. A graph-based network is defined as its underlying structure. Nodes may represent users, relying parties or attribute providers. The edges reflect confidence paths between consumers of attributes and attribute providers for particular properties. A service provider is able to decide, on the basis of aggregated confidence paths, on trust in a certain attribute that is offered by a specific attribute provider. The confidence on a path between consumer and attribute provider refers to the corresponding distance between the two nodes. In contrast to our model, the AttributeTrust framework uses the notion of confidence for a particular provider and issued property. There is no aggregation of trust or confidence for different attribute providers offering the same attribute. Furthermore, no accumulation of trust on the level of the digital identity is modelled. However, in our opinion the aggregation of trust on claims and digital identities is a significant prerequisite for an easy and automated trust decision without choosing and evaluating specific attribute providers.

III. Trust in Blockchain-based Identity Management

Trust is a very significant part of social relationships and everyday life with a wide variety of different meanings and definitions [7]. A person may trust another if she or he can rely on the correctness of the other person's statements or on their adherence to agreements. A company may be trusted by consumers based on the compliance of marketing promises with the reality of the product in the user's real-life experience. Trust can be based on previous experiences that shape the reputation of an entity or on recommendations of others that are directly trusted by the user. Due to the subjective and general nature of trust, that is applicable in various domains, a holistic and absolute definition is missing [8].

In our opinion, the general definition of Decision Trust, stated by Josang et al. [7], is the most applicable definition to characterise trust in blockchain-based identity management. Decision Trust is the degree a person deliberately goes into dependency to another entity in a particular circumstance. Nevertheless, a person's preference of assurance is still satisfied, despite that a negative impact may occur. In other words, a person willingly relies on somebody accepting adverse consequences in case the other person or organization violates a finalized, potential informal, agreement or provides false statements.

In terms of identity management, considering digital identities, claims and attestations, the service provider or any other relying party depends on the identity provider for correctness and validity of the provided information. A service provider needs to trust that the digital identity is valid. Furthermore, trust into claims is required to rely on correctness and actuality of the statements. Moreover, trust into attestation issuers to properly attest claims is an additional significant demand. Referring to the initially stated example on age verification, a service provider needs to trust the shown identity card as basis for a decision on age. Making a positive decision compels the provider to be dependent on the correctness of the information. In case false data is provided and the consumer is a minor, despite the age verification, a lawsuit with potential punishment could be the negative consequence. The required trust for a specific situation and information strongly depends on the extent of the potential negative consequences as well as the subjective risk appetite the service provider is willing to take. Risk might be the loss of financial gains or jurisdictional prosecution.

In addition to trust considerations on the overall identity management layer as application domain, the used blockchain technology requires reputation and trust management in additional functional components. The consensus protocol is applied by the nodes of the blockchain network in order to agree on the validity of the next block. Various security considerations in terms of attacks and defenses are researched [9]. Uncovering an attack may affect the reputation of a node in a negative way, finally reducing the trust in the node by the other nodes of the network. Besides the consensus protocol, the peer-to-peer communication tier is relevant for trust management. On this layer, numerous attacks, e.g. providing malformed data, denial-of-service against specific nodes, exists as well [10]. Degrading reputation based on metrics derived from misbehaviour forms a trust measure to increase resilience of the communication.

IV. A Quantifiable Trust Model

The following subsections outline our quantifiable trust model for blockchain-based identity management. Starting with objectives and requirements of the concept, we define afterwards a representation of the attestation, claims and digital identity model as a directed graph. Subsequently, we describe functions to calculate the specific trust values and present the corresponding algorithm. Additionally, we derive qualitative trust categories.

A. Objective

Our quantifiable trust model targets the identity management layer of a blockchain-based identity management system. The model's objective is to provide a simplified and automated way to reason about the trustworthiness of digital identities and their claims without evaluating single attestation issuers. On the one hand, service providers and relying parties are enabled to decide on their subjective layer of trust for providing a service. A dedicated analysis for trusting numerous attestation issuers is not anymore required. Therefore, SSI adoption is generally facilitated. On the other hand, trust values are usable to differentiate digital identities into two classes. They are separated into those that have a real-world subject and those that are created for other, potential malicious, purposes. This prevents an attacker from counterfeiting multiple identities and thwarts a Sybil attack [11]. Considering the blockchain network as a service provider, the digital identities characterized by a high trust value may serve as foundation for majority or vote-based blockchain consensus algorithms.

B. Requirements

Besides our objective, we consider the following requirements for designing the quantifiable trust model.

- Self-Policing [11]. The generation of trust values takes place in the distributed network without the support of a trusted third party. As an essential requirement, the attribute self-policing directly supports the vision of blockchain networks in remediating central governance bodies.
- No Benefit for New Entrants [11]. Newly created digital identities have no advantage over existing identities in terms of trust. Otherwise, it would be beneficial to replace digital identities and facilitate the whitewashing attack.
- Trust Decision on Attestation, Claim and Digital Identity. Service providers are enabled to make separate trust decisions on digital identity, claim and attestation level.

C. Design Decisions

The conceptual foundation of the quantitative trust model is aligned to the previously described objective and requirements. An identity is comprised of claims and attestations that are issued by other identities. In the formal representation of our trust model we consider identities, claims and attestations as distinct objects, although in general claims and attestations are an intrinsic component of an identity. The trust score of an attestation is directly derived from the issuing digital identity. A claim's trust value is determined from attestations issued to the claim. Eventually, the trust metric of a digital identity itself is deduced from claims that are issued to it and therefore, from the attestations of the respective claims. The concept is a closed circuit of trust between digital identities of the blockchain-based identity management system.

As the concept is a closed loop, initial trust is granted to pre-trusted digital identities. These initial trust holders spread the trust throughout the network by issuing attestations to other digital identities. A careful selection of the pre-trusted digital identities is fundamental for securing the concept.

Trust values of the different entities are projected into the interval $[0, 1)$. The value 0 denotes no trust at all. A result close to 1 indicates excellent trust. Figures in between reflect limited trust on various levels. We determined a fixed interval for trust values to enable an interpretation of a lower and upper border of trust. That allows a judgement in categories determining the trustworthiness of a digital identity, besides comparing the trust values of several digital identities among each other.

When setting the lower bound of the interval to 0, meaning no trust, we do not utilize the concept of distrust. Distrust could be modelled as a negative value, e.g. expanding the interval to a lower bound of -1. Nevertheless, on the one hand it fosters the whitewashing attack under the assumption that a new digital identity has no bias toward both trust and distrust maxima. Then the replacement of the old digital identity by a new identity with a 0 trust score is lucrative as long no mechanism prevents the creation of new identities for the same subject. On the other hand, a new identity starting with maximum distrust in a closed trust interval is comparable to the initial assignment of no trust.

We apply a quantitative model with continuous trust scores to facilitate the computation in a distributed execution environment as the blockchain network. Additionally, the quantitative scores enable fine granular and individual trust decisions by service providers. One service provider may require a higher score than another to consider a digital identity or a specific claim as trustworthy. Besides that, qualitative categories of trust can be deduced due to the fixed range of trust values.

D. General Model

We model the trust scheme in the form of a directed graph as a naturally fitting structure for the concept. The vertices V comprise digital identities, claims and attestations. The edges E represent the relations between the nodes. The sets E and V are finite, and their elements can be arbitrarily ordered and referenced by an index. Additionally, the graph is accompanied by three functions specifying the trust values for attestations T_A, claims T_C and digital identities T_I. The trust model TM is formally defined as follows.

$$TM = (V, E, T_A, T_C, T_I)$$

Subsequently, we define the elements of the TM trust model.

E. Vertices V

The set of vertices V is comprised of the sets of attestations A, claims C and digital identities I. It is formally modelled as the following.

$$V = I \cup C \cup A$$

The sets of digital identities, claims and attestations are disjoint.

$$I \cap C \cap A = \emptyset$$

A digital identity is an abstract object that is distinguished by an identifier. Additionally, the identifier is used to reference this object. The set of digital identities is comprised of all digital identities and is formally defined as follows.

$$I = \{i \mid i \text{ references a digital identity}\}$$

The set of claims contains expressed statements by identities. A claim is issued for an identity. Therefore, the existence of a relation between a claim and an identity is a prerequisite.

$$C = \{c \mid c \text{ is a claim } \wedge \exists i \in I : (c, i) \in E\}$$

The set of attestations includes released attestations. An attestation is issued to a claim by an identity. Therefore, the existence of a relation between an identity and a claim is a requirement.

$$A = \{a \mid a \text{ is an attestation} \\ \wedge (\exists c \in C : (a, c) \in E) \wedge (\exists i \in I : (i, a) \in E)\}$$

A trust value t is assigned to each vertex and determined by the respective trust function. We denote t_{a_j} to refer to the trust value of a specific attestation a with index j. Trust values for claims c and digital identities i are referenced alike using the specific labels.

F. Edges E

The edges connect different vertices and imply a trust flow from the originating to the receiving node. The set of edges E is formally defined as follows.

$$E = CI \cup AC \cup IA$$

Comparable to the definition of vertices, the distinct sets containing the edges are disjoint.

$$CI \cap AC \cap IA = \emptyset$$

The set CI is comprised of connections from claims to identities. Basically, it represents the relationship of a claim that is issued to an identity. A claim references exactly one identity. Additionally, a claim does not reference several digital

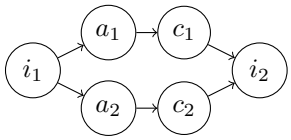

Fig. 1. Simple TM Graph

identities and there are not two similar claims assigned to the same digital identity.

$$CI = \{(c,i) \mid (c \in C) \wedge (i \in I) \wedge (\nexists (c,j) \in CI : j \in I \wedge i \neq j)\}$$

Attestations assigned to claims are contained in the set AC. Elements of AC reflect the association of an attestation with a claim. An attestation is created for exactly one claim. Furthermore, an attestation does not relate to several claims and the same attestation for the same claim does not exist several times.

$$AC = \{(a,c) \mid (a \in A) \wedge (c \in C) \wedge (\nexists (a,z) \in AC : z \in C \wedge c \neq z)\}$$

The set IA is comprised of connections from digital identities to attestations. In essence, a digital identity issues attestations for claims. An identity can create several attestations for different claims. However, exactly one attestation is created for one claim. Different digital identities do not issue the same attestation, but can attest the same claim.

$$IA = \{(i,a) \mid (i \in I) \wedge (a \in A) \wedge (\nexists (k,a) \in IA : k \in I \wedge k \neq i)\}$$

Fig. 1 presents a simple example of a TM graph. There are two identities: i_1 and i_2. Digital identity i_2 issued two attestations, a_1 and a_2, for separate claims, c_1 and c_2. These claims refer to identity i_2. In Fig. 2 a complex example of a TM graph is outlined. There are four identities, i_1, i_2, i_3, i_4, issuing several attestations to different claims. For instance, claim c_4 receives attestations a_4 and a_5. Additionally, claim c_2 acquires attestations a_2, a_3 and a_5. Taking this in consideration the model expresses a higher trust in claims c_2 and c_4 compared to claim c_3 with a single attestation under the assumption that the trust score of the attestations are similar.

Overall, the TM graph model represents the flow of trust between different elements. The graph does not show explicitly the relationship of an element issued by another object. The issuance relationship and the trust flow is the same for attestations. However, the 'issued by' relation is not represented for claims. The graph shows solely the trust flow. Claim c_2 references identity i_3 as attribute belonging to identity i_3. The graph does not show the identity that issued claim c_2. Deriving further from the outlined TM graph model, trust of an attestation flows directly to a claim and indirectly to a digital identity. Additionally, a digital identity is able to issue attestations to claims pointing to itself and therefore, create self-attestations. A claim that naturally target more than one identity can be represented in a divided manner as several claims referencing one identity each. Thus, opening the possibility to contradict mutually.

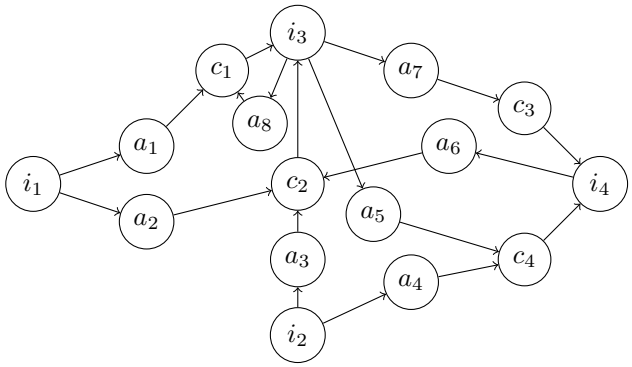

Fig. 2. Complex TM Graph

After defining the graph of the TM trust model, we outline the different functions to determine the trust flow from one node to another.

G. Attestation Trust Function T_A

According to the model, the trust of an attestation $a \in A$ for a claim $c \in C$ is solely determined by the trust of the digital identity $i \in I$ that issues the attestation. We formally define the attestation trust function T_A as follows with t_i denoting the trust value of identity i.

$$T_A : [0,1) \mapsto [0,1)$$

$$T_A : (t_i) \mapsto \begin{cases} t_i \cdot d, & t_i \geq \epsilon_{T_A} \wedge (c,i) \notin CI \\ 0, & t_i < \epsilon_{T_A} \vee (c,i) \in CI \end{cases}$$

The variable d is a discount factor to decrease the flow of trust from the digital identity to the attestation. The discount factor reflects declining trust of statements moving away from the actual source in social communities [8]. The element ϵ_{T_A} represents a lower threshold up to which the trust of the digital identity is further distributed to attestations. Discount factor d and ϵ_{T_A} facilitate the calculation of the algorithm in order to bring it to a final conclusion in a closed circuit system. We define both parameters as follows to achieve a smooth trust decrease.

$$d = 0.9 \quad \epsilon_{T_A} = 0.01$$

H. Claim Trust Function T_C

The trust of a claim is determined by the trust of the received attestations. An increased number of attestations indicates a higher trust value of the claim because several digital identities assert the validity of the claim. An additional factor that influences the trust in a claim is the trust value of each attestation. A large trust value of the attestation transfers more trust on the claim. Therefore, we define the claim trust function as follows.

$$T_C : [0,1) \times \ldots \times [0,1) \mapsto [0,1)$$

$$T_C : (t_{a_1}, \ldots, t_{a_n}) \mapsto 1 - e^{s \cdot ((\sum_{i=1}^{n} t_{a_i}) \cdot n)}$$

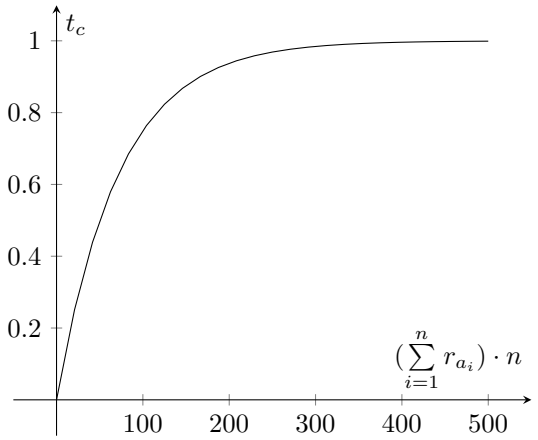

Fig. 3. Graph of T_C

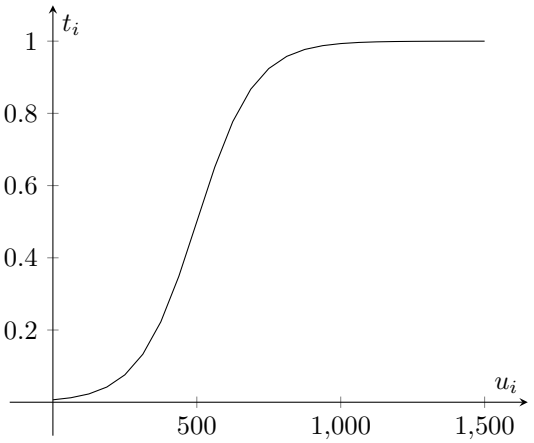

Fig. 4. Graph of T_I

The calculated trust values of function T_C are in the range from 0 to 1. No attestations result in 0 trust for the claim. In the positive direction, the function T_C converges to 1. The factor s shapes the slope of the approximation to 1. We chose the converging function to represent trust that initially aggregates fast and later on accumulates lower on a high level. This behaviour emulates a declining discriminability of high trust regions and denies an absolute trust interpretation.

The TM trust model in the present case is a web of trust between digital identities comparable to the Pretty Good Privacy (PGP) web of trust concept among key holders. A key holder signs keys of others to assert the ownership to a specific person. In 2012 a key in the PGP web of trust had on average of 10 signatures from other key holders [13]. Taking that into consideration, we decided to define the value of s in optimizing the function T_C towards a medium trust level of 0.5 by supplying 10 attestations with an average trust value of 0.5. Therefore, we determine s as the following value.

$$s = \frac{\ln(0.5)}{50}$$

In Fig. 3 the graph of the final function T_C is shown. On the x-axis the already aggregated trust values of the attestations multiplied with the quantity of the attestations is displayed. The y-axis outlines the resulting trust score for a claim.

I. Digital Identity Trust Function T_I

Trust in the digital identity is derived from the trust of the claims that are issued to the identity. A higher quantity of claims result in an increased trust value of the digital identity, because the identity is more realistically projected. The variety of attributes supports alignment to real-world persons that are comprised of a multitude of characteristics. Additionally, claims with high trust values lead to a larger increase of trust in the digital identity compared to claims with lower trust scores. Furthermore, the quantity of distinct attestation issuers of the claims influences the trust of the identity.

In this regard, we model first of all the claim trust accumulation function U_I as follows with u_i denoting the aggregated claim trust of identity i.

$$U_I : [0,1) \times \ldots \times [0,1) \mapsto \mathbb{R}^+$$

$$U_I : (t_{c_1}, \ldots, t_{c_n}) \mapsto (\sum_{i=1}^{n} t_{c_i}) \cdot n \cdot l$$

In function U_i the element l defines the number of distinct attestation issuers of the claims $c_1, \ldots, c_n$. Subsequently, we determine the identity trust function T_I.

$$T_I : \mathbb{R}^+ \mapsto [0,1)$$

$$T_I : (u_i) \mapsto \begin{cases} \frac{1-e^{k\cdot(u_i-f)}}{1+e^{k\cdot(u_i-f)}} \cdot 0.5 + 0.5, & u_i > 0 \\ 0, & u_i = 0 \end{cases}$$

The trust function T_i converges on the one side to 0 and with increasing u_i to 1. That behaviour creates an entrance barrier to reach a medium trust level to assert effort. Within a medium area of trust an increase and decrease in trust is accelerated due to the slope growth. It reflects immediate positive or negative changes on a medium trust level. In the high trust area the increase is again modelled evenly to represent decreasing ability for the differentiation of high trust values. The identity trust function T_I applies the parameters f and k. The constant f is determined to have an average trust of 0.5 by 10 claims with average trust of 0.5. Our reasoning is based comparably on the PGP web of trust rationale described in the previous section. Besides that, k is optimized to have a break below threshold $\epsilon_{T_I} = 0.01$ at the entry point of $u_i = 0$. Therefore, we determine f and k as follows.

$$f = 500 \quad k = -0.001$$

The characteristic graph of function T_i is presented in Fig. 4. On the x-axis the accumulated claim trust values for an identity is shown. The y-axis outlines the resulting trust score of the identity.

J. Algorithm

Updates on trust values of all vertices in our TM trust model are calculated in a round-based manner. This approach is aligned to the functioning of a blockchain network

that evolves with subsequent transactions. Initially, the trust scheme is comprised of pre-trusted digital identities with a particular initial trust score. Transactions may introduce additional identities or create claims as well as attestations. Adding identities or claims does not initiate a re-calculation of the trust values, because the trust flow in the model remain unchanged. Apart from that, appending an attestation modifies the trust flow and requires updates of trust values. The same situation applies in the case of removing identities, claims or attestations. Changes may lead to a propagation of subsequent updates on other objects in the TM trust model. The propagation of changes is finally completed based on the trust downslide incorporated into the T_A function.

K. Qualitative Trust Levels

Qualitative categories of assurance are easier to understand than quantitative values for judging trust by human beings [8]. Using the digital identity trust function T_I qualitative classes can be determined based on the quantitative trust score. We define the following levels exemplarily.

- No Trust ($0 \leq t_i \leq 0.2$). The digital identity is not trustworthy at all. Service provider may solely offer uncritical or public services to the user of the identity with the single purpose of recognising recurring users, for instance, to publish comments in blogs or in a forum.
- Limited Trust ($0.2 < t_i \leq 0.4$). The trustworthiness of the digital identity is restricted. Relying parties may solely accept the identity for uncritical services with the intention to raise the barrier for re-entry with a new identity. This is reasonable for the usage of reputation systems to minimize whitewashing after accumulation of negative feedback.
- Medium Trust ($0.4 < t_i \leq 0.6$). The digital identity has an average trust. Service provider may embrace the identity in situations having a minor risk. For instance, a person orders goods in a web shop for a limited two-digit amount. The failure of payment due to the abuse of the digital identity may be seen as bearable by the service provider.
- High Trust ($0.6 < t_i \leq 0.8$). The trustworthiness of the identity is high. Service provider may accept the identity in circumstances where a higher risk proposition exists. An example is the booking of hotel rooms or apartments.
- Superior Trust ($0.8 < t_i < 1$). The digital identity's trust is superior and can be accepted by service providers for highly critical applications or actions that are restricted by legal regulation. For instance, based on a superior trusted identity a bank account is opened.

As previously discussed trust between parties is a subjective matter. Besides the defined trust categories, the flexibility of the proposed model TM enables to determine distinct classes that may fit better the needs of service providers. Furthermore, trust decisions may not solely include the digital identity itself, but additionally incorporate a verification of the trust level for a specific claim that is significant for the offered service.

V. Threat Modelling

A beneficial objective for an attacker is to manipulate trust schemes to maliciously increase the trust score of a digital identity. Service providers rely on the trust values to offer their products to consumers. Attackers may carry out fraud requiring a trustworthy digital identity as foundation. In this section we present typical attack scenarios against trust schemes and evaluate the impact on our quantifiable TM trust model.

Self-attestations are attestations issued by a digital identity to claims referencing to oneself. The TM trust model allows this behaviour according to the formal description in order to reflect real-world situations. Persons make statements about themselves to introduce claims that are available for attestations of others. An attack vector is to directly influence the trust score of the claim, or indirectly the digital identity, in case self-attestations are included in to the trust computation. The presented TM trust model does not include self-attestations into trust calculations and is, therefore, not vulnerable to this threat.

Malicious groups of users are a comparable threat in a larger context. They exploit their digital identities to mutually issue attestations for invalid claims within their group to augment the respective trust scores. The attack scenarios impact is limited by two factors. On the one hand trust needs to flow into the group from an outside identity before it can be distributed within the group. The subject of the outside identity may revoke attestations of this trust flow in case abuse is uncovered. On the other hand issuing a high volume of attestations requires a comparable amount of transactions. Within a blockchain network each transaction has a particular processing fee. These payments restrict the attack probability as an additional factor.

The whitewashing attack describes the replacement of a digital identity with a new one to eliminate the old digital identity's negative trust history. In our TM trust model this attack vector does not apply as a digital identity starts with no benefit compared to existing identities. The trust value is 0 upon creation and cannot fall below later on.

Discrimination is an additional attack vector. A digital identity is discriminated by other identities. Therefore, it receives no attestations leading to the impracticality of increasing its own trust value. The higher the quantity of identities that belong to the discriminatory attacker group, the more difficult it is to find honest peers for receiving attestations. The discrimination attack may occur in the TM trust model. However, the pre-trusted peer groups cope with this attack as they are assumed to be honest. Furthermore, after a certain bootstrapping period and wide adoption of the blockchain-based digital identities it is unlikely that the attack is launched comprising a majority of the identities.

A traitor is a person who initially behaves honestly and receives attestations to build up trust. Having a high trust score the behaviour changes to malicious intent. The high trust score is used to facilitate trust increase in malign groups

or create attestations for invalid claims. The TM trust model discourages this attack based on the entry barrier for increased trust values on identity level. Building up initial trust is a certain effort that may not be put at risk to change to a malicious behaviour. Upon disclosing the malicious actions, attestations might be revoked by the respective issuer.

VI. Discussion

The described TM trust scheme creates a beneficial model to automate reasoning on the trustworthiness of claims and digital identities in blockchain-based identity management. However, the concept poses the following two major challenges.

The security assumptions of the TM trust model are obviously based on the pre-trusted group of digital identities as initial source of trust. This centralizes the decentralized nature of the blockchain-based trust model to a certain extent. Additionally, malicious members within this group are able to easier subvert the trust scheme in case of collusion compared to other groups that need to build up trust in a demanding way. To alleviate this weak spot, the pre-trusted peer group must be chosen carefully and distributed. Furthermore, a larger group size counteracts the centralization and collusion potential.

Another challenge in the current trust scheme is the non-existence of a punishment process to penalize malicious behaviour. An attestation of a claim leads to a trust increase on the claim and therefore, on the digital identity as claim holder. There is no immediate differentiation between a maliciously attested claim with invalid content and the attestation of a valid claim that leads to a justified trust flow. However, the disclosure of a false attestation causes its revocation and subsequently removes the granted partial trust. Despite that, there is no additional penalty on the attestation issuers trust score to reflect this misbehaviour and discourage similar actions in future.

VII. Future Work

An important enhancement is the incorporation of a dispute mechanism covering punishments for incorrect attested claims. This measure significantly strengthens security against certain threat models. In blockchain networks trust and reputation is additionally considered on the consensus and peer-to-peer communication layer. Incorporating and aggregating this information to the identity management layer on attestations, claims and identities seems to be beneficial for increasing the expressiveness of the respective trust scores.

In the model, initial trust is bootstrapped in the form of pre-trusted digital identities. Research on other forms of initial trust distribution is an interesting field to further remediate centralization.

An attestation of a claim creates a trust flow from one issuing digital identity to another. Thereby, the trust flow is independent from the context of the attested claim. As a way of increasing trust flow expressiveness, the integration of context dependent attestations is an interesting area.

VIII. Conclusion

Trust in blockchain-based identity management requires a different approach compared to trust in the traditional identity provider paradigm because of the decentralized nature. To address this challenge, we devised a novel quantifiable trust model with the additional benefit of automating and simplifying trust reasoning. Digital identities, claims, attestations, as well as their relations, are modelled as directed graph and trust functions that define the flow of trust from one element to another. There is no necessity for costly evaluation of each attestation issuer as the foundation for a trust decision. Based on a continuous trust scale for identities, we derived qualitative levels of trust for better reasoning on the expressed trust level for humans. Finally, we analysed our trust model with regard to popular attack patterns on trust schemes.

References

[1] S. Nakamoto, Bitcoin: A Peer-to-Peer Electronic Cash System. 2008.

[2] C. Meinel, T. Gayvoronskaya, M. Schnjakin, Blockchain: Hype oder Innovation. Hasso-Plattner Institut. 2018.

[3] C. Allen, The Path to Self-Sovereign Identity. 2016. Url: http://www.lifewithalacrity.com/2016/04/the-path-to-self-sovereign-identity.html

[4] A. Tobin, D. Reed, The Inevitable Rise of Self-Sovereign Identity. A white paper from the Sovrin Foundation. 2017.

[5] I. Thomas, C. Meinel, An Identity Provider to manage Reliable Digital Identities for SOA and the Web. In Proceedings of the 9th Symposium on Identity and Trust on the Internet. 2010.

[6] I. Thomas, C. Meinel, An attribute assurance framework to define and match trust in identity attributes. In Proceedings of 2011 IEEE International Conference on Web Services. 2011.

[7] A. Josang, R. Ismail, C. Boyd, A Survey of Trust and Reputation Systems for Online Service Provision. In Decision Support Systems. 2007.

[8] M. Tavakolifard, K. C. Almeroth, A Taxonomy to Express Open Challenges in Trust and Reputation Systems. Journal of Communications Volume 7. 2012.

[9] M. Conti, S. K. E, C. Lal, S. Ruj, IEEE A Survey on Security and Privacy Issues of Bitcoin. 2017.

[10] W. Lin, Attacks Against Peer-to-peer Networks and Countermeasures.

[11] J. R. Douceur, The Sybil Attack. In Proceedings of First International Workshop on Peer-to-Peer Systems. 2002.

[12] S. D. Kamvar, M. T. Schlosser, H. G. Molina, The EigenTrust Algorithm for Reputation Management in P2P Networks. In Proceedings of the 12th international conference on World Wide Web. 2003.

[13] H. P. Penning. Analysis of the strong set in the PGP web of trust. Url: https://pgp.cs.uu.nl/plot/

[14] D. Donato, M. Paniccia, M. Selis, C. Castillo, G. Cortese, S. Leonardi, New Metrics for Reputation Management in P2P Networks. In Proceedings of the 3rd International Workshop on Adversarial Information Retrieval on the Web. 2007.

[15] C. Ziegler, G. Lausen, Spreading Activation Models for Trust Propagation. IEEE International Conference on e-Technology, e-Commerce and e-Service. 2004.

[16] A. Singh, L. Lium, TrustMe: Anonymous Management of Trust Relationships in Decentralized P2P Systems. In Proceedings of the 3rd. International Conference on Peer-to-Peer Computing. 2003.

[17] I. Thomas, C. Meinel, Enhancing Claim-Based Identity Management by Adding a Credibility Level to the Notion of Claims. In Proceedings of the IEEE International Conference on Services Computing. 2009.

[18] A. Krishna, V. Varadharajan, A Hybrid Trust Model for Authorisation Using Trusted Platforms. In the Proceedings of International Joint Conference of IEEE TrustCom. 2011.

[19] A. Mohan, D. M. Blough, AttributeTrust - a Framework for Evaluating Trust in Aggregated Attributes via a Reputation System. In the Proceedings of the Sixth Annual Conference on Privacy, Security and Trust. 2008.