Towards Robust and Effective Trust Management for Security: A Survey

Dongxia Wang, Tim Muller, Yang Liu, and Jie Zhang
School of Computer Engineering
Nanyang Technological University
Singapore, Singapore
Email: wang0915@e.ntu.edu.sg, {tmuller,yangliu,zhangj}@ntu.edu.sg

***Abstract*—There is a need for robust and effective trust management. Different security problems result in different requirements to the design of trust management, and the existing attacks in trust management for security are yet to be solved. In this paper, we first propose a framework to classify desired properties of trust management for each type of security problems. We then investigate typical representative attacks and existing solutions in trust management for security. By considering both these security properties and attacks on trust management systems, our work serves to propel the design of more effective and robust trust management systems for security.**

I. Introduction

Traditional security mechanisms are faced with challenges, which can be elegantly addressed with trust management. Traditional mechanisms do not address security threats caused by internal malicious agents, while trust management systems can detect them by evaluating their trustworthiness or reputation. Further, traditional security mechanisms are not designed to protect against threats from malicious service providers [26], while trust management can be used to protect both service consumers and providers.

Although the purpose of different trust management based security approaches is similar – namely, to guide security decisions based on trust evaluation results – the requirements to make them effective vary with regard to different security problems and circumstances. For example, trust management based access control requires privacy preserving, service-orientation, and revocation of invalid on-going access. Therefore, it is necessary to find and summarise the critical issues in trust management design for each type of security problems – which we embark on in Section III.

Robustness of trust management is the other side of the coin. The importance of robustness is evidenced by the presence of attacks, such as the misleading feedback attack, the on-off attack, the discrimination attack and the Sybil attack. These attacks affect the accuracy of trust evaluation, negatively impacting security decisions based on these trust evaluations. Hence, defending against potential attacks should be another consideration of trust management design.

There are different classes of surveys regarding trust management. Some of them mainly discuss trust related issues like definitions of trust, properties of trust relationships, trust classification, trust formation, and various metrics of trust [55]. Some of them focus on trust management used to solve a specific kind of security problems, like routing issues [10, 14, 49]. Most of them discuss trust management based security in a particular environment, such as WSNs [17, 39, 53] MANETs [4], VANETs [48, 55], Multi-Agent systems [19], P2P systems [13], cloud computing [31], and web service [7, 26]. We provide a survey considering the requirements of different security problems for effective trust management, without overlooking the important issue of robustness in trust management.

In this paper, we first discuss challenges faced with traditional security approaches but can be solved by trust management (Section II). Then we study the recent trust models built for each kind of security problems, namely authentication, access control, secure service provision and secure routing (Section III). Based on these studies, we identify some desired properties of trust management design for each type of security problems. In addition, we inspect robustness issues of trust management (Section IV). We investigate potential attacks, all of which finally aim at trust evaluation, and some existing solutions. We conclude that there is no effective solution for several attacks like discrimination. Therefore we call upon more efforts towards the design of robust and effective trust management systems for security.

II. Challenges to Traditional Security Mechanisms

Traditional security deals with security guarantees with regard to security problems such as authentication and access control. These guarantees can be based on encryptions, certificates or credential verification. In open, distributed networks with more dynamic security challenges, these mechanisms may be inadequate to obtain security guarantees. Trust management may help mitigate security issues – without guarantees – by avoiding interactions with high security risks.

Whether agents actually are malicious is immaterial in traditional security; the important question is whether agents could act maliciously. Therefore, internal misbehaviour (e.g., selfish behaviour, malicious behaviour) is typically not detected by traditional mechanisms. Trust management is designed to be able to rate agents based on their behaviour records. Therefore, trust management can help avoid malicious agents.

Second, naive authentication and authorisation protocols are no longer effective in networks like MANETs. The premise of naive authentication is that the system knows the identities of legal agents in advance. This clashes with the premise of, e.g., pervasive computing, where devices are casually accessible and mobile. In some cases, the system cannot know in advance which devices are going to access, and what their access rights are [27]. Trust management based access control systems are designed to address this challenge.

Third, traditional security mechanisms are typically not designed to protect against threats from malicious service providers [26]. They typically aim to protect services or providers from malicious consumers. However, consumers also need to receive high quality services, while protecting their privacy from malicious providers. This can be referred to as secure service provision. For example, in mobile agent systems, agents need to be protected from malicious tampering of hosts. Trust management is effective at addressing these challenges, as detailed in the next section.

III. Trust Management for Security

Different types of security problems and circumstances have different requirements to the design of effective trust management. In this section, we discuss four categories of trust-based security solutions – authentication, access control, secure service provision and secure routing. For each, we identify the important requirements, and see whether solutions in the literature adhere to the requirements. Note that our requirements overlap with those in the traditional security perspective.

A. *Trust Management-based Authentication*

In authentication, one verifies that the identity of a person or object is what it claims to be [32]. Trust management is introduced to facilitate authentication in various applications [5, 8, 12, 44]. Here, we give a brief introduction to some of such approaches.

In [44], an authentication protocol is built to allow entities from one cluster to communicate with entities in another cluster. An agent that wants to communicate with the target agent in a new cluster needs to present certificates of its trust value, which are used for authentication. These certificates are signed by introducers in its original cluster.

In VANETs, where the number of authenticating executors is small compared to the number of on-board units (OBUs), OBUs need to wait for the nearest authenticating executor to authenticate before it can access services. TEAM [5] is designed to reduce the waiting time. An OBU is regarded as trusted after being authenticated successfully, and will be authorised to authenticate not-yet-trusted OBUs. This mechanism builds chains of transitive trust relationships rooted in authenticating executors, which speeds up authentication.

In a federated identity management system, an agent's authentication assertions can be created and propagated across different authorities. This requires service providers to evaluate the trustworthiness of the agent's identity. In [12], authentication trust of a agent is used to evaluate whether the identity is legitimate.

Agents may not want to authenticate by providing personal information to untrustworthy entities. Therefore, in [8], trust management happens before the authentication process. Only when trust evaluation result is higher than a certain threshold, the authentication phase will be started.

TABLE I
Properties of trust management based authentication models

	Park et al.	Gomi	Chuang et.al	ElHusseini et al.
	(2009) [44]	(2010) [12]	(2011) [5]	(2013)[8]
Mutual Authentication	Yes	Yes	N/A	N/A
Global Newcomer	Monitor	N/A	N/A	N/A
Privacy Preservation	N/A	Anonymity location privacy	N/A	Non-sensitive data used

Regarding these models, there are some key issues that are worthwhile to highlight and discuss below.

1) Mutual Authentication: In centralised systems, agents may have to unilaterally authenticate to a server. However, mutual authentication is vital in environments where two entities know little about each other, or where authentication protocols cannot always operate normally.

In [44], the target agent also needs to provide a certificate of its trust value to the requesting agent. In TEAM [5], when the authenticating executor authenticates an OBU, the OBU needs to ensure that the authenticating executor is genuine.

2) Global Newcomers: A global newcomer is an agent which is new to the whole network. It has no past interactions, which means there is no evidence to evaluate its trustworthiness. Global newcomers must be considered by the authentication mechanism.

In [44], a newcomer will first be monitored by all the other agents in a cluster for a certain time, based on which trust value will be computed. In [8], risk assessment (based on second-hand information) is used for dealing with agents which are not evaluated before. However, these agents are not global newcomers as we define. How global newcomers are treated is not specified.

3) Privacy Preservation: Protecting the privacy of agents being authenticated is important. Agents are reluctant to provide too much personal information for authentication. Authentication should avoid this.

Non-sensitive information is used in the evaluation of trust in [8]. TEAM satisfies anonymity and location privacy.

In conclusion, an effective trust management based authentication protocol should achieve mutual authentication, privacy preservation, and be able to deal with global newcomers. Table I summarises the properties of several recent trust management-based authentication models. It can be seen that none of the models are sufficiently effective in achieving the desired properties above.

B. *Trust Management-based Access Control*

In distributed networks where resources for each agent are limited (e.g., limited processing power, memory space, battery life and bandwidth), resource discovery is vital. Unconstrained resource discovery, however, may lead to security threats. Access control is needed to restrict unauthorised access to resources based on security policies of the system.

In highly dynamic networks like MANETs, traditional access control approaches which rely on identity (e.g., mandatory access control, role based access control) are not feasible. In trust based approaches, access rights are decided based on trust evaluation of the requesters and the security policies. Here, we discuss some key issues regarding trust based access control approaches we surveyed.

1) Service-Oriented Access Control: Not all services of a provider or device require the same level of security. For example, write access to a file may require a different security level than read access. Different security levels require different degrees of trust. Hence, a uniform trust threshold for all the services is infeasible. Thus, trust based access control should be service oriented, rather than device oriented.

In [33], accesses to services with different security levels are assigned different trust thresholds, allowing dubious data requestors with low trust values to access some low-risk services but not high-risk services.

2) Privacy Protection: Privacy protection is crucial when agents' personal information is being collected and used, especially in e-business environments.

In [33], an authorization of a data item depends on the requested time interval. Personal information is only kept for the period required to serve its purpose. In [2], the model allows data owners to control the degree of data disclosure according to its privacy level. Personal data items are classified into different privacy levels based on the privacy preference of the data owner. Data item with higher privacy levels will be kept for a shorter time period.

3) Continuity of Access Rights: Continuity means the presence of on-going access rights [34]. After access is granted, new requester events (e.g., malicious behaviours) and system attributes may be received by the access control manager. If these events indicate that the requester cannot be trusted anymore, on-going access should be revoked.

Re-calculation and re-evaluation systems are introduced in [34] and [33]. The re-calculation system is responsible for re-calculating the trust value of the requester based on new evidence received during on-going access. The re-evaluation system is used to check if the access control rules are violated. On-going access rights would be revoked if either of these two systems receives negative results.

4) Competitors' Recommendation: Providers of the same service are often competitors, trying to maximise their own revenue. As a result, they may be reluctant to warn each other about malicious requestors, or even provide dishonest recommendations. An effective trust evaluation method takes this into account, rather than blindly incorporating recommendations. In fact, this issue relates to the robustness property of trust management that is discussed in more details in Section IV.

In [16], providers are assumed to only be able to use their own data. But situations of inefficient first-hand experience are ignored in this case.

TABLE II
PROPERTIES OF TRUST MANAGEMENT BASED ACCESS CONTROL MODELS

	Li et al.	Li et al.	Gupta ea.	Bhatia
	(2009) [34]	(2011) [33]	(2011) [16]	(2013)[2]
Service Oriented	No	Different trust thresholds	No	Service specific
Privacy Protection	Yes	Yes	N/A	Yes
Continuity of Access Rights	Re-calculation Re-evaluation	Re-calculation Re-evaluation	N/A	Access rights revocation
Competitors' Recommendation	N/A	N/A	Rely on own data	N/A

In summary, for effective trust based access control, following properties are desired, namely service oriented access control, privacy preserving, revocation of invalid on-going access and filtering malicious recommendations from competitors. The properties of the recent access control models are listed in Table II. None of the models meet all of the requirements we discuss.

C. Trust Management-based Secure Service Provision

Authentication and access control protect service providers from malicious requesters. The reverse, protecting requesters from malicious service providers, however, can also be of importance [26]. Secure service provision exists in service provision networks where trust already plays a role (e.g., eBay or Amazon) and in networks where providers have more control over the data of the requesters (e.g. mobile agent systems and cloud computing environments).

Service provision networks like Amazon offer lots of open trading opportunities for consumers and providers, allowing providers to be malicious. In [40], trust relations are evaluated in a normal way. Providers are evaluated based on whether they honour the agreements built with consumers in their past performance. The approach combines first-hand experiences with second-hand evidence from other consumers.

In mobile agent systems, agents need to be protected from malicious execution hosts (execution service providers), which can cause agents' code to be disclosed, agents' data to be changed, and agents sent to wrong destinations. In MobileTrust [35], execution trust – which is the measure of trustworthiness of a host – is used to detect and eliminate malicious hosts.

In hybrid cloud computing environments, where both private and public clouds exist, the customers' control over their data is diminishing once their data is processed by third-party clouds [29]. In this situation, consumers need to ensure the trustworthiness of the cloud providers. In [1], a trust and reputation system is established to enable cloud customers to evaluate the trustworthiness of cloud providers, and select best cloud services.

In conclusion, these models all attempt to select a service provider by evaluating its trustworthiness or reputation. Unlike other security problems, privacy is the main concern of trust based secure service provision. Besides, service-contract consistency should also be considered in trust evaluation, to select both secure and high-quality services.

D. Trust Management-based Secure Routing

Secure routing is a routing technique in which the sender of a packet determines the complete sequence of agents through which to forward the packet [23]. Routing is vital for systems where agents cannot communicate directly to the destination agents (e.g., in MANETs and WSNs, agents can only communicate to neighbours within radio range). Message forwarding has to depend on collaborations among agents. Due to selfish or malicious intent, however, some agents may not collaborate as expected. Moreover, defective agents may also introduce faults. Both of these misbehaving agents threaten routing security.

The goal of trust based routing is to select trustworthy neighbours as packet forwarding candidates. Typically, each neighbour is assigned a routing score, and the agent with the highest score will be selected [3, 9, 18, 41, 50, 54]. Here, we do not detail each trust based routing model, rather, we discuss some key issues in the models. On the basis of this, we want to highlight the requirements for an effective trust based routing scheme.

1) Trust Metrics: Different trust metrics capture different aspects of security of routing. An optimal metric should consider all of the potential security threats in routing.

In the routing model THWMP [41], trust is simply based on packet loss. Each agent calculates the packet loss by its upstream neighbour, with which it updates trust value of the neighbour.

In most of the models, however, there is a collection of trust metrics. In ATSR [54], eight trust metrics are combined, while each one stands for an aspect of security concern, such as forwarding (to detect agents denying to forward packets), packet precision (to ensure that no unexpected modification has occurred). These metrics are based on detectable events and can be used to measure them inversely. For instance, packet modification can be measured by the packet precision metric. By these metrics, trust evaluation can well capture various types of misbehaviour, which can then help improve the ability to resist them. In both [3] and [18], direct trust evaluation depends on the percentage of successful interactions, which in [18], is defined as forwarding the message to the correct peer, but in MTR [3], is not defined. Although it is claimed in [3] that an agent's trust is based on quality of service characteristics, such as packet forward and data rate, there is no explanation how they are implied in defining successful interactions.

2) Trust Evidence Propagation: In networks where agents are highly mobile, such as MANETs, VANETs, and CNRs, the neighbours of an agent change frequently, which causes it to have a smaller number of interactions with a larger number of partners [52]. As a result, there are not enough experiences for an agent to evaluate arbitrary partners. Therefore, effective trust evaluation should be based on both direct experiences and indirect trust evidence.

All of the surveyed models incorporate indirect trust evidence into calculation.

3) Routing Score: For security, trust value should be a factor of routing score which is used to select the next-hop. At the same time, it would be better if routing distance is incorporated, which impacts routing efficiency.

THWMP [41] decides wether to add agents to a path solely based on trust evaluation results. The remaining models all consider both trust value and the distance to destination, for the purpose of selecting trustworthy agents with less physical latency to the destination.

ATSR [54], takes remaining energy in the agent into consideration. Regardless of computation complexity, models considering distance would be more efficient in packet forwarding. In each model, a weighted sum function is proposed to aggregate these metrics.

DTEGR [50] optimises the static weighting scheme in ATSR. It selects agents with trust values above a threshold to form a forwarding list, from which the agent with the closest distance to the destination will be chosen as the next hop.

In conclusion, there are three requirements for effective trust based routing. First, trust evaluation should capture as much potential misbehavior as possible. Second, indirect trust evidence should be incorporated (correctly) when the direct experiences are not enough for trust evaluation. Third, functional requirements on the routes should be considered, and a balance must be achieved between secure routes and efficient routes.

The properties of the models are provided in Table III. All models have some trust metrics, in various degrees of detail. Trust evidence propagation is present in all models, albeit implemented differently. Except THWMP, all models use both trust values and routing distance in the score.

E. Discussion

Different security problems have some common requirements on trust management (e.g., privacy protection). We will discuss these in detail below. Further, we compare trust-based security mechanisms to traditional approaches.

1) Common Requirements: We identify three common requirements:

- *Privacy protection* is a consideration in all of security problems above. Authentication should avoid requiring private information. In access control, data owners need to specify security policies for data of different privacy levels, which should be combined with trust decisions. Protection of consumers' data is also an important component of service provision trust. In secure routing, data (including personal information) is often encrypted, to prevent internal agents from snooping.
- *Trust evidence propagation* is desired in environments where first-hand experience is insufficient to make effective trust decisions. In the aforementioned security problems, this is the case; most prominently in secure routing (Section III-D). The requirement for trust evidence propagation, however, depends on the characteristics of the

TABLE III
PROPERTIES OF TRUST MANAGEMENT BASED SECURE ROUTING MODELS

	THWMP (2013) [41]	ATSR (2013) [54]	DTEGR (2013) [50]	MTR (2013)[3]	Fenye et al. (2012) [9]	Han et al. (2013) [18]
Trust Metric	Packet loss	8 types[1]	8 types[1]	4 types[2]	QoS trust and social trust	Packet precision
Trust Evidence Propagation	Yes	Yes	Yes	Yes	Yes	Yes
Routing Score	Trust value (TV)	TV, distance	TV, distance	TV, distance	TV, distance	TV, distance

[1] Packet precision, network ACK, forwarding, confidentiality, authentication, reputation response, reputation validation, remaining energy.

[2] Packet precision, data rate, reliability, power consumption.

environment. Competitors' recommendations may be dishonest and should be inspected (Section III-B). Generally, this is a misleading feedback attack (Section IV-B) which can come from either service providers or requesters. Regardless of the type of security problems, it should be considered in trust management where second-hand evidence is used for trust evaluation.

- *Global newcomers* should explicitly be taken into consideration, in trust management systems where they occur.

2) Difference with the Traditional Security: Being social control mechanisms, trust-based security schemes are unlike traditional security mechanisms. The former can be regarded as *soft security* approaches, while the latter can be regarded as *hard security* approaches [45]. Hard security strives to guarantee that secure components work as intended. However, it is not feasible to guarantee security of all components in all systems. Without trust management, the system would be left unprotected. Soft security acknowledges the existence of malicious entities and behaviours, and it attempts to detect them and accordingly decreases the impact caused by them. Additionally, in traditional security, there are typically no security levels, just secure or not secure – hence the term *hard* security. In trust management, trust evaluation provides a quantitative value for the object, which can represent various levels of security.

3) Combine with the Traditional Mechanisms: Trust management can be combined with traditional mechanisms to support security. Trust management evaluates entities based on their behaviours, while traditional security relies on rigourous mechanisms (e.g., certificates, credentials). In [35], the evidence results from these two are combined to make security decisions. There are models in which trust management is combined with role based access control [11, 46]. Specifically, in [11], access is granted if both a client's trust level exceeds a threshold and the global role and permissions are correct.

IV. ROBUST TRUST MANAGEMENT

Trust management helps to identify trustworthy entities as secure. However, to maximise profit, malicious entities may strategically attack trust management systems. For example, malicious entities may provide dishonest ratings trying to defame an honest agent. A weak trust management system may not function as desired under these attacks. Hence, robustness is crucial in the design of trust management for security.

A. The Role of Robustness in Trust Management

The accuracy of trust evaluation is closely related to the robustness of trust management systems, which can further impact trust-based security decision making. Jøsang and Golbeck state that the correctness of the computed trust score is influenced by two factors: robustness of trust systems, and attack incentives [25]. The lack of incentives can reduce the number of attacks, and more robust systems can mitigate these attacks, both leading to increased accuracy. It has been shown that attacks exist in current trust management systems [28]. A robust trust management system is a system where there are less attacks, or where the attacks' effectiveness is limited. For accurate trust evaluation when applying trust management to support security, we cannot ignore the robustness.

B. Attacks and Solutions

In order to study robustness, we must study potential attacks. We study the typical attacks. Attacks result from agents with malicious intent. We cannot detect or prevent malicious intents. We can, however, mitigate the damage caused by malicious behaviours – behaviour resulting from malicious intents – or disincentivise attacks. For example, we can detect unfair ratings and filter them out or detect the dishonest raters first and abandon their ratings [56]. When the attack relies on the vulnerabilities of the system, then the system must be fixed. For example, if the system assumes one account per agent, then it is crucial for the authentication system to identify multiple identities registered by one agent. Otherwise the so-called Sybil attack or the newcomer attack (see below) can happen. Although a comprehensive solution against all of existing attacks does not (yet) exist, researchers have proposed methods to mitigate some of them.

1) Misleading Feedback Attack: The misleading feedback attack is also referred to as the unfair rating attack or the badmouthing attack. It results from dishonest recommenders who attempt to corrupt the reputation of good entities, or increase the reputation of their conspirators. There are three kinds of schemes to defend this attack: recommender trust-based schemes [36, 47, 56], detection-based schemes [20, 37] and incentive and punishment-based schemes [57]. In the first type, recommender trust, which represents trust to a recommender's honesty in providing second-hand evidence, is used to filter recommendations. Detection-based schemes apply data mining approaches such as clustering and classification [20, 37]. Incentive and punishment-based schemes aim to reward truthful feedback to reduce the occurrence of fake recommendations.

For recommender trust-based schemes, Sun et al. propose a strategy which is characterised as follows [47]: Recommendations from raters with a high recommender-trust value are more capable of propagating. Recommendations from those with lower recommender-trust have smaller impact on decision-making. In [36], raters' trust values, which are used to weight their ratings, are derived from both local and global rating information (ratings about other sellers). The reputation of the seller is derived by aggregating weighted ratings.

For detection-based schemes, both [20] and [37] apply a clustering method to identify dishonest raters. Raters are clustered, where rating differences act as the distance measure. Raters that are in the same cluster as the buyer are regarded as honest, since they have smaller rating difference, and thus similar rating behaviours. Ratings with multiple levels are considered in [37], where rating vectors are constructed using the number of transactions rated with a level as a component. Multi-criteria ratings are considered in [20], where different clusters are formed for different sets of criteria.

For incentive-based schemes, in [58], truthful feedback is encouraged by making sellers provide increased quality of products with decreased prices to those reputable buyers. Each buyer keeps a group of advisors, consisting of the most trusted fellow buyers. Sellers identify reputable buyers based on the number of advisor groups they belong to. Honest buyers will benefit from its ratings by gaining more profitable transactions. In [38], a limited inventory of each seller is considered, where buyers compete with each other to get the purchase. In a naive system, buyers would provide negative feedback about high quality sellers, since they are scarce. They propose an incentive mechanism where buyers providing truthful ratings are assigned higher score, which makes them have more chance to transact with reputable sellers.

2) Discrimination Attack: In a discrimination attack, the service provider provides high quality services to some groups, but low quality services to others. This induces contradictory ratings among these groups, which may impact their trust value as recommenders. If a group identifies dishonest recommenders based on rating difference to its own (like using the detection-based schemes in Section IV-B1), then the group which provides contradictory ratings will be regarded as dishonest. To defend this attack, self-experiences should not be set as the only benchmark to identify dishonest recommenders. We found no effective solution for this attack.

3) On-off Attack: On-off attack means malicious entities behave inconsistently over time, exploiting the trust computation algorithm, while remaining undetected [47]. For example, an agent firstly accumulates a high trustworthiness through good behaviour. Then, additional ratings play a smaller role in changing its reputation, and it starts behaving badly while maintaining an acceptable reputation. This suggests that older behaviour records may indicate less about an agent's current behaviour.

To address this problem, the most commonly used approach is to introduce a forgetting factor [47]. However, a fixed forgetting factor can also be used by malicious entities to facilitate the on-off attack. With a long forgetting factor, the computed trust value does not reflect the current state of the agent, whereas with a short forgetting factor, the behaviours are forgotten quickly, and the agent regains its trust too easily. Sun et al. propose an adaptive forgetting scheme [47]. When the trust value is below the threshold, a longer forgetting factor is used, otherwise, a shorter forgetting factor will be used. Therefore, the trust value can keep up with the change in the agent's behaviours, and moreover, recovery from a low trust value requires enough good behaviours.

In P2P systems, the on-off attack is called dynamic personality of peers. Xiong and Liu [51] propose an adaptive time window-based algorithm to react to such personalities. The idea is to adaptively choose a smaller time window to collect the most recent behaviour records of a peer, when its performance drops. The trust value computed from those most recent records will be compared with the one computed from all records in a larger time window. If it is lower than a certain threshold, which indicates the peer is performing badly recently, then it will be set as the peer's trust value.

4) Sybil Attack: The Sybil attack comes from malicious entities who freely create several identities. The attacker can use different identities each time to behave maliciously, and then the blame will be shared by all of these identities, instead of being afforded by itself. Also, relying on its multiple identities, the attacker can give multiple ratings over the same service object, unfairly increasing its influence on the service's reputation. Countermeasures against Sybil attacks are usually confined to a particular network (e.g., VANETs [15], P2P [6], WSNs [30]). In [43], admission control is used to block unnecessary raters when there is enough information to predict the rating value of a service item. Based on this intuition, only ratings from the reliable raters will be used for prediction of the rating value.

5) Newcomer Attack: An agent may cause a newcomer attack if it can easily register a new identity. By re-registering, the attacker can easily get rid of its previous bad behaviour history, and bad reputation. The newcomer attack is also called the re-entry attack [25]. Similar countermeasures as against the Sybil attack may work here. In addition, a penalty for new agents works effectively against newcomer attacks (however, punishing new agents may be unacceptable in many settings).

6) Value Imbalance Exploitation: Typically, ratings do not indicate the value of the services. A malicious agent can gain high profits and also reputation by providing more high quality services with low value, while providing low quality services with high value. To defend this, one simple way is to assign weights to ratings as a function of the value of services [25].

C. Discussion

1) Attack Model: In this section we summarise some attacks and existing solutions in trust management systems. These attacks happen in different stages of a trust management system: the Sybil attack and newcomer attack happen in

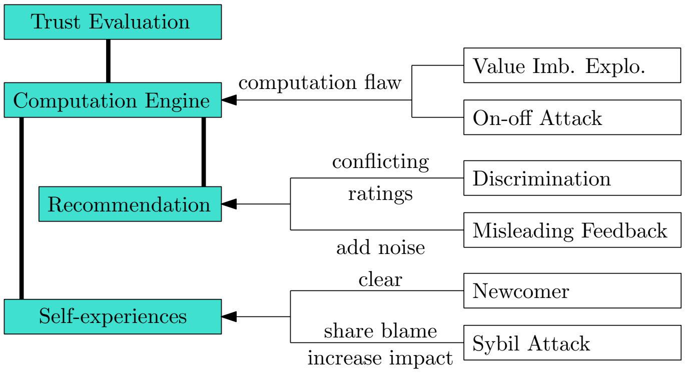

Fig. 1. Attack Model

the login phase; the discrimination attack, the on-off attack and value imbalance exploitation happen in the transaction phase; and the misleading feedback attack happens in the trust evaluation phase. Jøsang proposes a concrete model for attack functional phases and attack vectors in trust and reputation systems [24]. Although occurring in different stages, we conclude that these attacks ultimately aim at trust evaluation. Their intentions and targets are presented in Figure 1.

First, value imbalance exploitation and on-off attacks actually attempt to exploit flaws on computation algorithms, to behave maliciously without proportionate negative impact on their reputation. In the former, when computing trust values, if the computation algorithm does not weight ratings for service of different values, then attackers can gain unfairly high profits while maintaining a disproportionately good reputation. In the latter, if the computation algorithm does not correctly weight old and recent behaviour, then attackers can adaptively oscillate between good and bad behaviours, and remain undetected.

Second, attacks such as misleading feedback and discrimination aim at second-hand evidence, thus impacting trust evaluation results. A misleading feedback attack consists of an agent providing dishonest ratings. Discrimination itself seems to have no direct harms (Section IV-B), however, discriminating groups will cause contradiction among ratings of these groups, which can influence their recommendation trust in each other.

Third, newcomer and Sybil attacks can impact both first-hand and second-hand evidence. Newcomer attackers attempt to get rid of its bad behaviour history, which may be first-hand or second-hand evidence to others. Sybil attackers can provide multiple ratings over the same service object and/or they diminish blame for their bad behaviour by spreading it over fake identities.

2) Future Work: Regarding robust trust management, there are many issues remaining to be solved. We distinguish three major areas for improvement:

First, some attacks, such as discrimination, proliferation and countermeasures, reputation lags, and exit attack, are underexposed in the literature. We need to study the extent of these attacks. Second, solutions for misleading feedback attacks in multi-agent based e-marketplaces are the most studied (can be seen in [20–22, 36–38, 56]). However, robustness regarding other attacks and systems are still underdeveloped, and some trust models ignore them. Third, the design of robust trust management can be more organised with general evaluation measures. First, we require a formal definition of robustness for trust management (as, e.g., in [42]). Then, there should be automated verification mechanisms for the robustness, as currently exist for program correctness or computer security. Finally, the evaluation measures should be applicable to all the existing vulnerabilities and attacks in trust management.

V. Conclusion

In this paper, we studied the effectiveness and robustness of trust management solutions for security. We first explored and presented what desired properties effective trust management design should have for each type of security problems. Then, we investigated robustness issues of existing trust management, which is also crucial for trust management design.

The study of effective trust management is based on following security problems: authentication, access control, secure service provision, and secure routing. For each of these problems, we identified crucial requirements they impose on trust management systems. These requirements generally differ between the problems. However, we found that some properties do not depend on the type of security problems. First, privacy protection is required in all of these problems. Second, second hand trust evidence should be used where self-experience is insufficient – where scrutiny is applied to second hand evidence.

Then, to study the robustness of trust management for security, we inspected existing attacks, which can not only impact the accuracy of trust evaluation, but also leave the system insecure. We found that all the attacks we studied eventually aim at trust evaluation, although they happen in different phases. We also found that there is a lack of effective solutions for attacks like discrimination. Finally, we call for general evaluation measures which can make robust trust management design more organised.

In all, the survey identifies requirements proposed by each type of security problems to effective trust management design. These desired properties, together with discussion on attacks and solutions, serve to propel the design of more robust and effective trust management for security.

Acknowledgement

This work is supported by the “Formal Verification on Cloud” project under Grant No: M4081155.020 and the “Verification of Security Protocol Implementations” project under Grant No: M4080996.020, and partially supported by the A*STAR SERC grant (1224104047) awarded to Dr. Jie Zhang..

References

[1] J. Abawajy. Establishing trust in hybrid cloud computing environments. In *Proceedings of the 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pages 118–125. IEEE, 2011.

[2] R. Bhatia and M. Singh. Trust based privacy preserving access control in web services paradigm. In *Proceedings of the 2nd International Conference on Advanced Computing, Networking and Security (ADCONS)*, pages 243–246. IEEE, 2013.

[3] Z. Chen, R. Zhang, L. Ju, and W. Wang. Multivalued trust routing based on topology level for wireless sensor networks. In *Proceedings of the 12th International*

Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pages 1516–1521. IEEE, 2013.

[4] J.-H. Cho, A. Swami, and R. Chen. A survey on trust management for mobile ad hoc networks. *IEEE Communications Surveys and Tutorials*, 13(4):562–583, 2011.

[5] M.-C. Chuang and J.-F. Lee. TEAM: Trust-extended authentication mechanism for vehicular ad hoc networks. In *Proceedings of International Conference on Consumer Electronics, Communications and Networks (CECNet)*, pages 1758–1761. IEEE, 2011.

[6] W. L. da Costa Cordeiro, F. R. Santos, G. H. Mauch, M. P. Barcelos, and L. P. Gaspary. Identity management based on adaptive puzzles to protect p2p systems from sybil attacks. *Computer Networks*, 56(11):2569–2589, 2012.

[7] N. Dragoni. A survey on trust-based web service provision approaches. In *Proceedings of the 3nd International conference on Dependability (DEPEND)*, pages 83–91. ACM, 2010.

[8] A. El Husseini, A. M'hamed, B. El Hassan, and M. Mokhtari. Trust-based authentication scheme with user rating for low-resource devices in smart environments. *Personal Ubiquitous Comput*, 17(5):1013–1023, 2013.

[9] B. Fenye, C. Ing-Ray, C. MoonJeong, and J.-H. Cho. Hierarchical trust management for wireless sensor networks and its applications to trust-based routing and intrusion detection. *IEEE Transactions on Network and Service Management*, 9(2):169–183, 2012.

[10] K. Garg, M. Misra, et al. Trust based security in manet routing protocols: a survey. In *Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India*, page 47. ACM, 2010.

[11] C. Ghali, A. Chehab, and A. Kayssi. Catrac: Context-aware trust-and role-based access control for composite web services. In *Proceedings of 10th International Conference on Computer and Information Technology (CIT)*, pages 1085–1089. IEEE, 2010.

[12] H. Gomi. An authentication trust metric for federated identity management systems. In *Proceedings of the 6th International Conference on Security and Trust Management*, STM'10, pages 116–131. Springer-Verlag, 2011.

[13] S.-F. Gong and Z. Jian-Lei. A survey of reputation and trust mechanism in peer-to-peer network. In *Proceedings of the International Conference on Industrial Control and Electronics Engineering (ICICEE)*, pages 116–119. IEEE, 2012.

[14] J. Gonzalez, M. Anwar, and J. Joshi. Trust-based approaches to solve routing issues in ad-hoc wireless networks: A survey. In *Proceedings of the International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pages 556–563. IEEE, 2011.

[15] J. Grover, M. S. Gaur, and V. Laxmi. A novel defense mechanism against sybil attacks in vanet. In *Proceedings of the 3rd international conference on Security of information and networks*, pages 249–255. ACM, 2010.

[16] B. Gupta, H. Kaur, N. Namita, and P. Bedi. Trust based access control for grid resources. In *Proceedings of the International Conference on Communication Systems and Network Technologies (CSNT)*, pages 678–682. IEEE, 2011.

[17] G. Han, J. Jiang, L. Shu, J. Niu, and H.-C. Chao. Management and applications of trust in wireless sensor networks: A survey. *Journal of Computer and System Sciences*, 80(3):602–617, 2014.

[18] Y. Han, K. Koyanagi, T. Tsuchiya, T. Miyosawa, and H. Hirose. A trust-based routing strategy in structured p2p overlay networks. In *Proceedings of the International Conference on Information Networking (ICOIN)*, pages 77–82. IEEE, 2013.

[19] Y. Han, S. Zhiqi, C. Leung, M. Chunyan, and V. Lesser. A survey of multi-agent trust management systems. *Access*, 1:35–50, 2013.

[20] A. A. Irissappane, S. Jiang, and J. Zhang. A biclustering-based approach to filter dishonest advisors in multi-criteria e-marketplaces. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 1385–1386, 2014.

[21] S. Jiang. Towards the design of robust trust and reputation systems. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 3225–3226. AAAI Press, 2013.

[22] S. Jiang, J. Zhang, and Y.-S. Ong. An evolutionary model for constructing robust trust networks. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pages 813–820, 2013.

[23] D. B. Johnson and D. A. Maltz. Dynamic source routing in ad hoc wireless networks. In *Mobile computing*, pages 153–181. Springer, 1996.

[24] A. Jøsang. Robustness of trust and reputation systems. In *Proceedings of the 4th International Conference on Self-Adaptive and Self-Organizing Systems Workshop (SASOW)*, pages 159–159. IEEE, 2010.

[25] A. Jøsang and J. Golbeck. Challenges for robust trust and reputation systems. In *Proceedings of the 5th International Workshop on Security and Trust Management (SMT)*, 2009.

[26] A. Jøsang, R. Ismail, and C. Boyd. A survey of trust and reputation systems for online service provision. *Decision support systems*, 43(2):618–644, 2007.

[27] L. Kagal, T. Finin, and A. Joshi. Trust-based security in pervasive computing environments. *Computer*, 34(12):154–157, 2001.

[28] R. Kerr and R. Cohen. Smart cheaters do prosper: defeating trust and reputation systems. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 993–1000, 2009.

[29] K. M. Khan and Q. Malluhi. Establishing trust in cloud computing. *IT professional*, 12(5):20–27, 2010.

[30] M. Klonowski and M. Koza. Countermeasures against sybil attacks in wsn based on proofs-of-work. In *Proceedings of the 6th International conference on Security and privacy in wireless and mobile networks*, pages 179–184. ACM, 2013.

[31] V. Kumar, B. Chejerla, S. Madria, and M. Mohania. A survey of trust and trust management in cloud computing. *Managing Trust in Cyberspace*, page 41, 2013.

[32] P. Lamsal. Understanding trust and security. *Department of Computer Science, University of Helsinki, Finland*, 2001.

[33] M. Li, X. Sun, H. Wang, Y. Zhang, and J. Zhang. Privacy-aware access control with trust management in web service. *World Wide Web*, 14(4):407–430, 2011.

[34] M. Li, H. Wang, and D. Ross. Trust-based access control for privacy protection in collaborative environment. In *Proceedings of the International Conference on e-Business Engineering (ICEBE)*, pages 425–430. IEEE, 2009.

[35] C. Lin and V. Varadharajan. Mobiletrust: a trust enhanced security architecture for mobile agent systems. *International Journal of Information Security*, 9(3):153–178, 2010.

[36] S. Liu, A. C. Kot, C. Miao, and Y.-L. Theng. A dempster-shafer theory based witness trustworthiness model to cope with unfair ratings in e-marketplace. In *Proceedings of the 14th Annual International Conference on Electronic Commerce*, pages 99–106. ACM, 2012.

[37] S. Liu, J. Zhang, C. Miao, Y.-L. Theng, and A. C. Kot. iclub: an integrated clustering-based approach to improve the robustness of reputation systems. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*, pages 1151–1152, 2011.

[38] Y. Liu and J. Zhang. An incentive mechanism designed for e-marketplaces with limited inventory. *Electronic Commerce Research and Applications*, 2013.

[39] J. Lopez, R. Roman, I. Agudo, and C. Fernandez-Gago. Trust management systems for wireless sensor networks: Best practices. *Computer Communications*, 33(9):1086–1093, 2010.

[40] M. Louta and A. Michalas. Trust management framework for efficient service provisioning in dynamic distributed computing environments. In *Proceedings of the Third International Conference on Internet and Web Applications and Services*, ICIW '08, pages 518–523. IEEE Computer Society, 2008.

[41] R. Mahajan, S. Singh, A. K. Bhardwaj, and P. Sharma. Trust based routing for secure wireless networking solutions. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3:14, 2013.

[42] T. Muller, Y. Liu, S. Mauw, and J. Zhang. On robustness of trust systems. In *Proceedings of the 8th International Conference on Trust management (IFIPTM)*, 2014.

[43] G. Noh, Y.-m. Kang, H. Oh, and C.-k. Kim. Robust sybil attack defense with information level in online recommender systems. *Expert Systems with Applications*, 41(4):1781–1791, 2014.

[44] S.-S. Park, J.-H. Lee, and T.-M. Chung. Authentication scheme based on trust and clustering using fuzzy control in wireless ad-hoc networks. In *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA)*, ICCSA, pages 345–360. Springer-Verlag, 2009.

[45] L. Rasmusson and S. Jansson. Simulated social control for secure internet commerce. In *Proceedings of the workshop on New security paradigms (NSPW)*, pages 18–25. ACM, 1996.

[46] I. Ray, D. Mulamba, I. Ray, and K. J. Han. A model for trust-based access control and delegation in mobile clouds. In *Data and Applications Security and Privacy XXVII*, pages 242–257. Springer, 2013.

[47] Y. L. Sun, Z. Han, W. Yu, and K. R. Liu. A trust evaluation framework in distributed networks: Vulnerability analysis and defense against attacks. In *Proceedings of the 25th International Conference on Computer Communications*, pages 1–13. IEEE, 2006.

[48] S. Tangade and S. Manvi. A survey on attacks, security and trust management solutions in vanets. In *Proceedings of the International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, pages 1–6. IEEE, 2013.

[49] U. Venkanna and R. L. Velusamy. Black hole attack and their counter measure based on trust management in manet: A survey. In *Proceedings of the 3rd International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom)*, pages 232–236, 2011.

[50] M. Xiang. *Trust-based energy aware geographical routing for smart grid communications networks*. PhD thesis, AUT University, 2013.

[51] L. Xiong and L. Liu. Peertrust: Supporting reputation-based trust for peer-to-peer electronic communities. *Trans. Knowledge Data Eng*, 16(7):843–857, 2004.

[52] H. Yu, Z. Shen, C. Miao, C. Leung, and D. Niyato. A survey of trust and reputation management systems in wireless communications. *Proceedings of the IEEE*, 98(10):1755–1772, 2010.

[53] Y. Yu, K. Li, W. Zhou, and P. Li. Trust mechanisms in wireless sensor networks: Attack analysis and countermeasures. *Journal of Network and Computer Applications*, 35(3):867–880, 2012.

[54] T. Zahariadis, P. Trakadas, H. C. Leligou, S. Maniatis, and P. Karkazis. A novel trust-aware geographical routing scheme for wireless sensor networks. *Wireless personal communications*, 69(2):805–826, 2013.

[55] J. Zhang. A survey on trust management for vanets. In *Proceedings of the International Conference on Advanced Information Networking and Applications (AINA)*, pages 105–112. IEEE, 2011.

[56] J. Zhang and R. Cohen. Evaluating the trustworthiness of advice about seller agents in e-marketplaces: A personalized approach. *Electronic Commerce Research and Applications*, 7(3):330–340, 2008.

[57] J. Zhang, R. Cohen, and K. Larson. A trust-based incentive mechanism for e-marketplaces. In *Trust in Agent Societies*, pages 135–161. Springer, 2008.

[58] J. Zhang, R. Cohen, and K. Larson. Combining trust modeling and mechanism design for promoting honesty in e-marketplaces. *Computational Intelligence*, 28(4):549–578, 2012.