A Survey of Privacy Preserving Reputation Systems

Omar Hasan

A Survey of Privacy Preserving Reputation Systems*

Omar Hasan
U. of Lyon, CNRS, INSA-Lyon, LIRIS, UMR5205, F-69621, France

Abstract

Reputation systems make the users of a distributed application accountable for their behavior. The reputation of a user is computed as an aggregate of the feedback provided by other users in the system. Truthful feedback is clearly a prerequisite for computing a reputation score that accurately represents the behavior of a user. However, it has been observed that users often hesitate in providing truthful feedback, mainly due to the fear of retaliation. Privacy preserving reputation systems enable users to provide feedback in a private and thus uninhibited manner. In this paper, we describe analysis frameworks for reputation systems and privacy preserving reputation systems. We use these analysis frameworks to review and compare the existing privacy preserving reputation systems in the literature. We identify the strengths and weaknesses of the various systems. We also discuss some open challenges.

1 Introduction

In recent years, reputation systems have gained popularity as a solution for securing distributed applications from misuse by dishonest users. A reputation system computes the reputation score of a user as an aggregate of the feedback provided by fellow users. Good behavior is rewarded by positive feedback and consequently a high reputation score. On the contrary, bad behavior results in negative feedback and a low reputation score, which can lead to isolation or exclusion from the application. Some examples of applications of reputation systems are as follows:

- According to a survey on fraud in e-commerce [19], fraud accounted for a total loss of US$ 2.7 billion in the United States and Canada

*Work in progress. Draft version 0.11.

in 2010. Reputation systems used by e-commerce websites (such as `ebay.com`, `amazon.com`) mitigate the risk that a seller would turn out to be fraudulent.

- Several cases have been reported where fake online persona have hijacked the identity of professionals and then succeeded in connecting to their real network of acquaintances [1]. Reputation systems that root out fake profiles on social networks include Unvarnished [60] and Duedil [26].

- There is a risk in peer-to-peer file sharing networks (such as BitTorrent) that a file uploaded by a seeder is fake. Reputation systems for defeating fake content in peer-to-peer file sharing networks have been proposed by Costa and Almeida [18], Yu [65], and Kamvar et al. (EigenTrust) [40].

- Nodes in Mobile Ad-hoc Networks (MANETs) depend on neighbors to route their messages. However, neighbors may be selfish and may drop messages to conserve their resources. Reputation systems for discouraging selfish behavior in mobile ad-hoc networks include those by Hu and Burmester [37], and Buchegger et al. [11, 10].

Reputation score is an aggregate of the feedback, therefore an accurate reputation score is possible only if the feedback is accurate. However, it has been observed that the users of a reputation system often avoid providing truthful feedback [55]. This is particularly true about negative feedback. The reasons for such behavior include fear of retaliation from the target entity or mutual understanding that a feedback value would be reciprocated.

A solution to the problem of lack of truthful feedback is computing reputation scores in a privacy preserving manner. A privacy preserving protocol for computing reputation scores operates such that the individual feedback of any user is not revealed. The implication is that the feedback provider is rendered uninhibited to provide truthful feedback.

In this paper, we describe analysis frameworks for reputation systems and privacy preserving reputation systems. We use these analysis frameworks to review and compare the existing privacy preserving reputation systems in the literature. Consequently, we identify the strengths and weaknesses of the various systems.

2 An Analysis Framework for Reputation Systems

In this section, we develop an analysis framework that identifies the various dimensions of reputation systems. Since privacy preserving reputation systems are fundamentally reputation systems, we can use this framework to analyze their non-privacy features. The analysis framework for issues specific to privacy is presented in Section 3.

Some fundamental concepts in reputation systems are as follows:

Source User A user u is said to be a source user of a user t if u has feedback about t in a given context. A source user can also be referred to as a *rater*.

Target User When a source user assigns feedback to a user t, or a user q initiates a query to determine the reputation of user t, the user t is referred to as the target user. We can also refer to the user t as the *ratee*.

Querying User When a user q initiates a query to determine the reputation of a user t, the user q is referred to as the querying user. A querying user can also be called the *inquirer*.

Reputation The reputation of a target user is any function that aggregates the feedback of its source users.

The analysis framework for reputation systems is graphically represented in Figure 1. In the following sections, we present the various dimensions of reputation systems.

2.1 Architecture

The architecture of a reputation system is one of the key factors in determining how the following activities are conducted:

- Feedback collection
- Feedback aggregation (reputation computation)
- Reputation dissemination

The three architectures are: centralized, decentralized, hybrid.

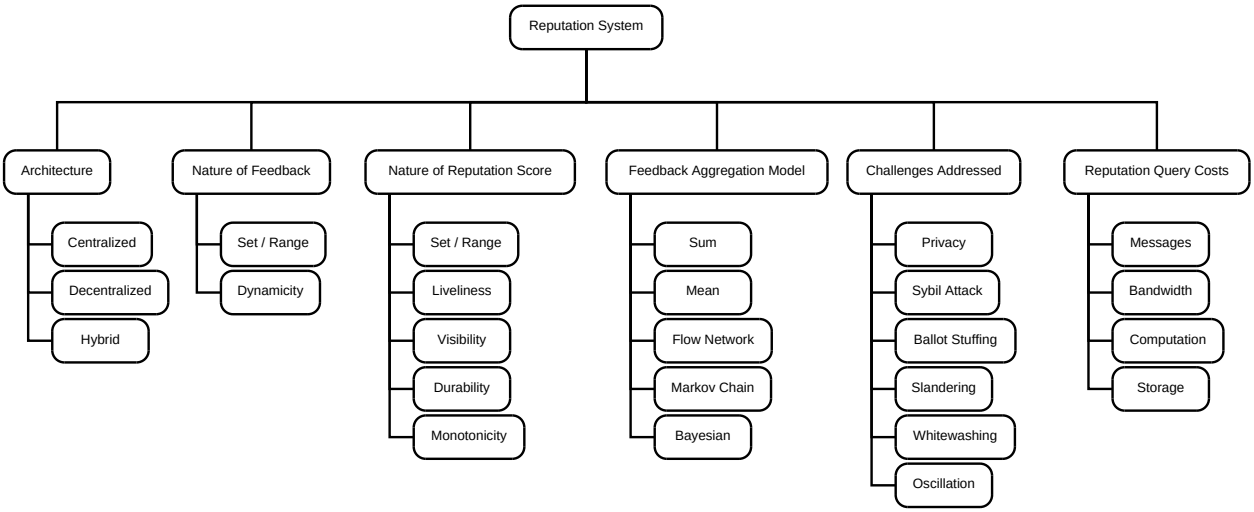

Figure 1: Analysis framework for reputation systems.

Centralized Centralized reputation systems are characterized by the existence of a trusted central authority. The central authority receives feedback from users, aggregates it to compute the reputation, and disseminates the reputation scores. One of the benefits of a centralized solution is that it is straightforward to implement. Additionally, the central authority is universally trusted, therefore users can be assured that the feedback collection, aggregation, and dissemination are being done correctly. However, if the central authority fails or becomes compromised, then the whole reputation system is compromised. Thus the central authority is a single point of failure and a high-value target for attackers. Centralized reputation systems are also unable to cater for decentralized environments such as peer-to-peer networks, ad-hoc networks, decentralized social networks, etc. Examples of centralized reputation systems include ebay.com, epinions.com, amazon.com, advogato.org, and PageRank [49].

Decentralized Decentralized reputation systems are suitable for decentralized environments as they do not assume the presence of a central entity. In decentralized reputation systems, a central location for submitting and aggregating feedback, and disseminating reputation does not exist. Feedback is commonly stored locally by the node who generates it, for example in response to his experiences with another party. Computing reputation of an entity in the system requires finding all or

a portion of the nodes who carry feedback about that entity. Once the feedback providers have been located, the aggregation may be done at a single location after receiving all feedback, or a more sophisticated protocol may be employed to aggregate the feedback in a distributed manner. Examples of decentralized reputation systems include Damiani et al. [20], Gupta et al. [32], EigenTrust [40], and PowerTrust [66].

Hybrid The hybrid architecture merges elements from the centralized and the decentralized architectures. Some activities are carried out in a centralized manner whereas others in a decentralized fashion. For example, in the reputation system by Androulaki et al. [3], *reputation coins* that represent feedback can be exchanged between users in a peer-to-peer manner. However, the reputation coins must be retrieved and deposited at a central entity called the *bank*.

2.2 Properties of Feedback

Feedback Set / Range The set or range that the feedback belongs to, for example, $\{-1, 0, 1\}$, $[0, 1]$.

Feedback Frequency A rater may provide separate feedback for each transaction with the ratee or one feedback value that reflects the cumulative experience of the rater with the ratee.

Feedback Dynamicity Feedback can be dynamic or not. Dynamic feedback can be altered with the passage of time and with altering experiences between rater and ratee. Whereas, non-dynamic feedback once provided remains constant. Additional feedback may be provided for new transactions, however, the earlier feedback cannot be altered.

Feedback Durability Feedback durability refers to the lifetime of a feedback value. A feedback value may remain valid for an indefinite period of time or it may be considered obsolete with the passage of time. An obsolete feedback value may be entirely excluded from the reputation computation or its significance may be discounted.

2.3 Properties of Reputation

Reputation Set / Range The set or range that the reputation belongs to, for example, $\mathbb{R}$, $[0, 1]$.

Reputation Liveliness As noted by Schiffner et al. [56] Reputation liveliness implies that a reputation system should not offer users the possibility to reach a final state of reputation in which bad behavior no longer damages their reputation. Thus, reputation should always consider all recent interactions or give users an indication that there are no more.

Reputation Visibility The visibility of a reputation score may be global or local. Global visibility implies that all nodes in the system view the same reputation score of a certain entity. Whereas with local visibility, the reputation score available to a subset of the nodes may be different than elsewhere in the system. Local visibility is generally a concern in decentralized reputation systems, where a different subset of feedback providers may be included for computing the reputation of an entity at different instances.

Reputation Durability Reputation durability refers to the transience of a reputation score. Once a reputation score is computed, it may be stored permanently for subsequent access by nodes through a simple retrieval operation. Recalculation of the score is mandated only when new feedback becomes available. Alternatively, the reputation score may be transient and re-computed every time a node wishes to learn the score. The latter approach requires repeated computation of the reputation, however, it does not require storage of the scores by a trustworthy entity.

Monotonicity Monotonic reputation implies that the reputation score increments in only one direction. For example, consider a reputation system in which a user can receive integer feedback between 1 and 5 for each transaction and reputation is considered as the sum of feedback. The reputation in such a reputation system can only increase upwards. The reputation of a user cannot be decremented.

2.4 Feedback Aggregation Models

There are a number of models for aggregating feedback to obtain reputation scores. We describe some of the common models below. A comprehensive survey of feedback aggregation models (also called reputation computation engines) is provided by Jøsang et al. [39].

Sum and Mean Model One of the most common methods of aggregating feedback to obtain the reputation score is simple summation. The

eBay reputation system (ebay.com) allows users to give positive (+1), neutral (0), or negative (−1) feedback. The reputation is computed as the sum of the feedback provided over a certain period of time. The reputation of a user is considered as high as the sum of the feedback. The advantage of this approach is that it is very straightforward and easy to understand for the users of the reputation system. A related method is to compute the reputation score as the mean of the feedback values. Reputation represented as mean has the benefit of being normalized and thus the reputation of different users may be compared objectively.

Flow Network Model A class of reputation systems (such as the Advogato (advogato.org) [44] reputation system) are constructed using the concept of flow networks. The users are considered as the nodes of a network and the feedback that they assign each other is considered as the flow in the network. The reputation of a node is computed as a function of the flow that the node receives from other nodes. A salient characteristic of such reputation systems is that a node cannot assign more flow to other nodes than it has received itself. This prevents a node from creating multiple pseudonyms for malicious purposes, since the total incoming flow and hence the reputation of the pseudonyms would be only as high as the original node itself. It is assumed in the Advogato reputation system that the amount of flow available in the network is constant and regulated by trustworthy nodes adjacent to the source.

Markov Chain Model Several reputation systems (such as EigenTrust [40] and PowerTrust [66]) draw on the Markov chain theory. Feedback from one node to another is considered as the probability of transition from the source to the target node. The reputation of a node is computed as the probability of arriving at that node by following random transitions from a known trustworthy node. The reputation systems based on the Markov chain theory also offer the advantage that a malicious node does not benefit from creating multiple pseudonyms for malicious purposes. This is due to the fact that even if the malicious node assigns maximum feedback to each of its pseudonyms, the probability of reaching those pseudonymous nodes from a trustworthy node would be no higher than reaching the original malicious node.

Bayesian Model The reputation score in a Bayesian reputation system is generally represented by a beta distribution in which the two free

parameters α and β correspond to the number of positive and negative feedback respectively. The reputation score is computed by statistically updating the given beta distribution. Bayesian reputation systems provide a sound mathematical basis for computing reputation scores [39]. Moreover, the observable difference in the statistical properties of fair and unfair ratings enables filtering out unfair ratings [63]. Bayesian reputation systems include those by Haller [33], Wang and Vassileva [62], and Jøsang and Ismail [38].

2.5 Challenges faced by Reputation Systems

Reputation systems can be classified by the challenges that they address and their success in resolving them. In this section, we discuss some of the challenges other than privacy that reputation systems have to contend with.

We list these challenges because when we review existing privacy preserving reputation systems, we would also like to analyze whether they address these problems in addition to the problem of privacy.

Sybil Attack The sybil attack [25] on a reputation system operates as follows: An attacker creates multiple identities in the system in order to gain an unfair advantage over honest users who own a single identity. The attacker may use its multiple identities to mount attacks including self-promotion, slandering, and ballot stuffing. The Advogato [44] and Appleseed [67] reputation systems prevent this attack by reducing the influence of pseudonyms created by a single entity. Since an entity has a limited amount of flow received from existing entities, it does not help to create new pseudonyms and distribute that flow among them. The total influence of the entity remains the same. Yu et al. [64] propose an approach based on social networks to detect sybil attacks. The algorithm operates by ensuring that the size of the cut between the set of known honest nodes and the set of potential attackers remains small.

Self-Promotion, Ballot Stuffing Self-promotion is the act of raising one's own reputation through unfair means. Self-promotion may be carried out by a user individually or in collusion with other members of the system. A self-promotion attack is possible in systems (such as eBay) where users may assign each other additional feedback after every transaction. Two users may repeatedly transact with each other, and after each transaction assign each other positive feedback. This attack is also known as ballot stuffing, which implies that a user

submits more feedback than he is entitled to. Another scenario is that a user creates multiple identities in the system (the sybil attack), and uses those fake multiple identities for self-promotion. The strategy employed by the reputation systems of many online auction and e-commerce websites (for example, eBay, Amazon), is to charge the seller a fee for each transaction. Thus, repeated fake transactions for the purpose of accumulating feedback becomes costly.

Slandering, Bad-Mouthing Slandering or bad-mouthing is the act of sabotaging an honest user's reputation by assigning them unwarranted low feedback. Motivation for such an attack may include retaliation, reducing a competitor's reputation, or malicious disruption of services. A slandering attack is particularly detrimental to the target user in applications that are sensitive to the presence of even a small amount of low feedback, such as high-value monetary systems. The reputation system by Belenkiy et al. [7] ensures fair exchange of feedback and services. The reputation system, oriented for content distribution peer-to-peer systems, uses cryptographic techniques to guarantee that when a peer receives the requested data blocks, he must provide positive feedback to the sender in return. Otherwise the data blocks stay locked and their content remains inaccessible to the peer.

Whitewashing A whitewashing attack occurs when a user with negative reputation quits the system and re-enters with a new identity and thus a fresh reputation. A reputation system is vulnerable to the whitewashing attack when: the pseudonyms in the system are not linked to real world identities, quitting the system incurs little or no loss, and creating new pseudonyms is cheap (in terms of limited resources, such as money, human effort, etc.). To mitigate the risk of whitewashing attacks, a reputation system may differentiate users who are newcomers from those who have been in the system for a long time. A user may only be allowed to build his reputation gradually by demonstrating good behavior consistently over a long period of time. This approach lessens the appeal of a whitewashing attack, since a user who re-enters the system with a new identity is not viewed as trustworthy. Systems that propose this approach include [34, 46].

Oscillation In oscillation, an attacker initially builds good reputation in the system and then suddenly shifts behavior to take advantage of honest users who are misled into trusting the attacker due to the good reputation. This attack is advantageous only if the payoff of the attack

is greater than the cost of building good reputation. One scenario is that an attacker engages in several low value transactions to accumulate reputation and then reverses its good behavior for a high value transaction. A reputation system may mitigate the risk of oscillation attacks by weighing feedback according to its age in the system. Systems that follow this strategy include Aringhieri et al. [6], Buchegger et al. [11], TrustGuard [57], and the Beta reputation system [38]. Swaminathan et al. [59] address this problem by setting a sales limit on each seller, which is bounded by the sum of the transaction costs (fees, insurance, shipping, etc.) paid by the seller thus far in the system. The seller may only sell items within its sales limit. The idea is that even if the seller suddenly shifts behavior and defrauds a buyer, he would not make any profit due to his past expenses.

2.6 Costs

The following two operations are executed in a reputation system: 1) querying the reputation of a target agent, and 2) housekeeping. The costs of each of these operations can be measured as follows:

1. Number of messages exchanged
2. Bandwidth consumed
3. Computational resources consumed
4. Storage required

Comparing these costs allows us to compare the efficiency of different reputation systems.

3 An Analysis Framework for Privacy Preserving Reputation Systems

In this section, we identify and explain the common dimensions and requirements of privacy preserving reputation systems. Developing this analysis framework of privacy preserving reputation systems would help us gain greater insight into prior research. It would also enable us to compare the different privacy preserving reputation systems in a normalized manner.

In Section 4, we identify the security objectives of privacy preserving reputation systems proposed in the literature. In section 5, we identify the

building blocks that serve as the foundation for privacy preserving reputation systems.

The analysis framework for privacy preserving reputation systems is graphically represented in Figure 2. It extends the framework for reputation systems presented in Figure 1.

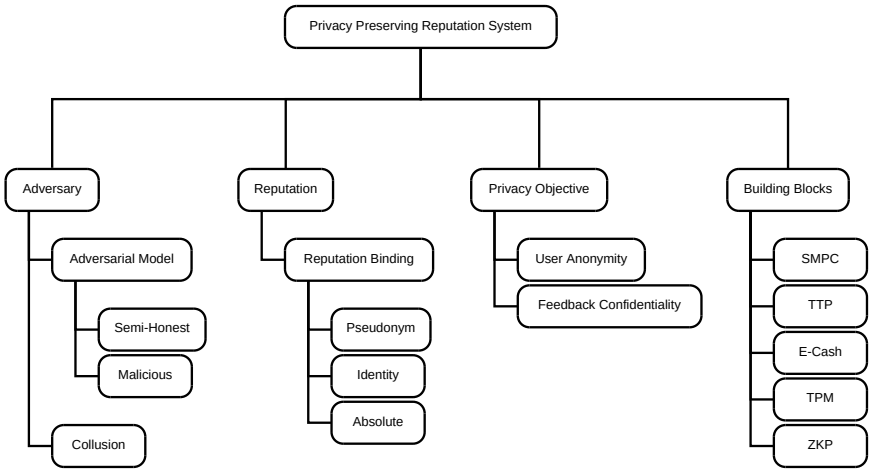

Figure 2: Analysis framework for privacy preserving reputation systems.

3.1 Adversary

The goal of a reputation system is to compute the reputation from the inputs of the participants. All participants of the protocol are expected to pursue this and only this goal. An honest participant is one who conforms to this expectation. However, there may exist dishonest participants who have ulterior motives. Those motives may include learning the inputs of other participants, tampering with the output, disrupting the protocol, etc.

3.1.1 Adversarial models

We identify three adversarial models, which characterize the behavior of dishonest users. The models are: Semi-Honest, Non-Disruptive Malicious, and Disruptive Malicious. A privacy preserving reputation system is considered secure under one of these models if it can show correctness and meet its privacy requirements under the given model.

Semi-Honest In the semi-honest model, the users do not deviate from the specified protocol. In other words, they always execute the protocol

according to the specifications. The adversary abstains from wiretapping and tampering of the communication channels. However, within these constraints, the adversary passively attempts to learn the inputs of honest users by using intermediate information received during the protocol and any other information that it can gain through other legitimate means.

Non-Disruptive Malicious Malicious users are not bound to conform to the protocol. Users under a malicious model may deviate from the protocol as and when they deem necessary. They actively attempt to achieve their objectives. They may participate in extra-protocol activities, devise sophisticated strategies, and exhibit arbitrary behavior. Specifically, malicious users may 1) refuse to participate in the protocol, 2) provide out of range values as their inputs, 3) selectively drop messages that they are supposed to send, 4) prematurely abort the protocol, 5) distort information, and 6) wiretap and tamper with all communication channels.

We define a non-disruptive malicious adversary as an adversary who executes the malicious actions only if they lead to the disclosure of the inputs of honest users. Non-disruptive users have a single objective: learn the inputs of honest users. They do not disrupt the normal function of the protocol other than to achieve this objective.

Disruptive Malicious We define a disruptive malicious adversary as an adversary who has the following objectives: 1) learn the inputs of honest users, and 2) disrupt the protocol for honest users. The reasons for disrupting the protocol may range from gaining illegitimate advantage over honest users to completely denying the service of the protocol to honest users.

3.1.2 Collusion

A dishonest user may act alone or multiple dishonest users may act in agreement to achieve their ulterior motives. When multiple dishonest users work together, it is referred to as collusion. Privacy preserving reputation systems either consider that collusion can take place between users or consider that collusion does not take place.

Collusion can be bounded or unbounded. Bounded collusion implies that the number of dishonest participants in the system allowed to collude with each other is limited, for example, 1/2 or 1/3 of all n participants.

Unbounded collusion places no limit on the number of dishonest participants who can collude with each other, thus $n-1$ of the participants can be dishonest and colluding, except for one honest participant whose privacy needs to be preserved.

3.2 Properties of Reputation

Privacy preserving reputation systems add the following property to the reputation score (in addition to the properties discussed for general reputation systems in the previous section):

3.2.1 Reputation binding

A privacy preserving reputation system can be either pseudonym bound or identity bound.

In a pseudonym bound system, the reputation of the user is associated with his pseudonym. If he changes or creates a new pseudonym then he looses his reputation. This can be disadvantageous for several reasons. This implies that reputation is not transferable between a user's multiple pseudonyms. Moreover, a dishonest user can drop a pseudonym with bad reputation and re-enter the system with a new pseudonym and a fresh reputation.

On the other hand, in an identity bound system, the reputation of a user is bound to his real identity. Even if he changes pseudonyms, he maintains his reputation.

4 Security Objectives of Privacy Preserving Reputation Systems

We have identified two broad categories of privacy preserving reputation systems with respect to their security objectives. The goal of the systems in the first category is to preserve the anonymity of the users. The systems in the second category do not hide the identity of the users but focus on preserving the confidentiality of the feedback that the users provide. The two categories of privacy preserving reputation systems are described as follows:

1. **Privacy preserving reputation systems with user anonymity.** The true identity of the users is hidden in these systems. The feedback providers thus remain anonymous. A user is represented in the system

by one or more pseudonyms which are unlinkable to his real identity. This framework allows the user to anonymously carry out transactions with others and submit feedback. There is no need to guard the confidentiality of the submitted feedback since the anonymity of the users prevents it from being linked to them.

2. **Privacy preserving reputation systems with feedback confidentiality.** These systems do not attempt to hide the identity of the users beyond assigning each user a single pseudonym. Moreover, these systems do not conceal the act of a user assigning feedback to another user. However, the value of the submitted feedback and any other related information is considered private. This type of systems is necessary since complete anonymity is not always possible due to the nature of real world transactions. For example, even if anonymity is preserved online on an auction site such as eBay, the exchange of physical items sold and bought through the site would reveal the real identities of the participants. Preserving the confidentiality of the feedback values is a practical alternative to enable users to submit truthful feedback without the fear of retaliation.

The security objectives of a privacy preserving reputation system can be further subdivided as those fulfilling privacy and those fulfilling correctness. The privacy objectives are concerned with hiding information about users, for example, preserving the anonymity of the rater and the ratee. On the other hand, the correctness objectives are aimed towards maintaining the integrity of the functions of the reputation system while preserving the privacy of the users. For example, correctness objectives include preventing a malicious user from manipulating the reputation aggregation function to forge an unmerited good reputation.

Figure 3 illustrates the classification of the objectives of privacy preserving reputation systems. In Sections 4.1 and 4.2, we describe several individual security objectives of privacy preserving reputation systems with user anonymity and with feedback confidentiality respectively. A particular reputation system may pursue a few or more of these objectives depending on the stringency of its security requirements.

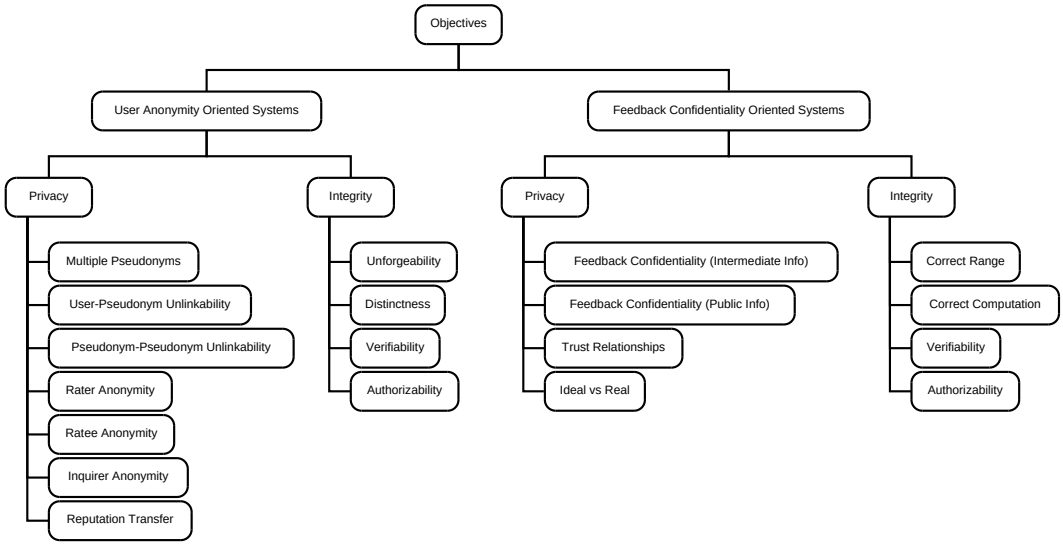

Figure 3: Objectives of privacy preserving reputation systems.

4.1 Privacy Preserving Reputation Systems with User Anonymity

4.1.1 Privacy Objectives

Multiple Pseudonyms A user is able to assume multiple pseudonyms in the system. As noted by Anwar and Greer [4, 5], the variation in the pseudonyms of a user may be on a per context or a per transaction basis. In the first case, a user may adopt a different pseudonym for each context in the system, for example, a tutor could use different pseudonyms for different subjects in an e-learning system. Alternatively, a user may choose a different pseudonym for each transaction in the system. The use of multiple pseudonyms makes the various contexts or the transactions of a user unlinkable to each other.

User-Pseudonym Unlinkability User-pseudonym unlinkability implies that the true identity of a user is not linkable to any pseudonym that he uses in the system. Androulaki et al. [3] identify this requirement as follows: Given a pseudonym P that does not belong to a corrupted party, the adversary can learn which peer owns P no better than guessing at random among all non-corrupted peers that appear consistent with P.

Pseudonym-Pseudonym Unlinkability Pseudonym-pseudonym unlink-

ability implies that two different pseudonyms that belong to the same user cannot be linked to each other. The adversary is unable to tell whether two given pseudonyms belong to the same user. This property is specified by Androulaki et al. [3] as follows: Given two pseudonyms P_1, P_2 that do not belong to corrupted parties, the adversary has no advantage in telling whether P_1, P_2 belong to the same peer or not. This requirement should hold as long as there are at least two non-corrupted peers who appear consistent with both P_1 and P_2 (because if there is only one such uncorrupted peer, clearly both pseudonyms belong to the same one).

Rater Anonymity A user is able to rate another user without his true identity being revealed. The purpose of rating anonymously is to prevent the adversary from linking the rater to his interaction with the ratee and the rating that he submitted. Schiffner et al. [56] state this property as follows: A pseudonym P_1 that interacted with a ratee R should not be linkable to the pseudonym P_2 that rated R.

Ratee Anonymity A user is able to receive a rating without his real identity being disclosed. A ratee may not wish to be associated with his past transactions and ratings since they could influence the ratings for his future transactions. According to Schiffner et al. [56], this property implies that a ratee R can use a different pseudonym for each transaction.

Inquirer Anonymity A user is able to inquire about the reputation of another user, however, others are not able to learn whose reputation he is querying or even the fact that he is inquiring about another user's reputation. Users wish to query the reputation of other users anonymously in order to prevent the adversary from compiling a profile of their interactions and interests.

Reputation Transfer and Aggregation A user is able to transfer reputation among multiple pseudonyms that he owns without letting anyone draw associations among these pseudonyms. Consequently, a user is able to aggregate the reputation of his multiple pseudonyms into the reputation of one pseudonym.

Unobservability of Reputation Transfer and Aggregation An adversary is unable to link the various pseudonyms of a user when he transfers reputation between them or aggregates their reputation.

4.1.2 Correctness Objectives

Reputation Unforgeability A user is unable to show reputation higher than the cumulative reputation of his pseudonyms. A user is unable to borrow good reputation from another user.

Distinctness It is possible to prove that the reputation of a target user is an aggregate of votes or feedback that come from distinct users while simultaneously hiding the identities of those users. The advantage of this property is that one or a few dishonest users are not able to submit multiple votes or feedback (ballot stuffing) for artificially raising the reputation of the target user.

Linkability in Case of Adversarial Behavior If and only if a user commits a predefined adversarial act such as ballot stuffing, then his pseudonym becomes linkable to his real identity. This property ensures that even though users are anonymous they are still accountable for any adversarial actions.

Exculpability The adversary is unable to frame an honest user for malicious behavior, such as ballot stuffing.

4.2 Objectives of Privacy Preserving Reputation Systems with Feedback Confidentiality

4.2.1 Privacy Objectives

Confidentiality of Feedback, with no Inference from Intermediate Information This property requires that a rating assigned by a rater to a ratee is never revealed to any other party including the ratee. The system must protect the confidentiality of the feedback such that the feedback is neither divulged explicitly nor inferred from any intermediate information gained by the adversary during a reputation query. The system may define the confidentiality of the feedback as deterministic or probabilistic. In the first case, the adversary is unable to learn any information about the feedback. However, in the case of probabilistic confidentiality, the amount of information leakage depends on certain variables, such as the number of raters, the reputation score, etc.

Confidentiality of Feedback, with no Inference from Public Information The reputation score of any user is by definition public and any other user in the system is authorized to learn this score. The issue is that

a dishonest user may use this public information to derive the private feedback of honest users. For example, in an additive reputation system, the adversary simply needs to observe the reputation score before and after the latest user submits his feedback to learn its value. The requirement of confidentiality of feedback, with no inference from public information implies that the adversary is unable to learn information about the feedback even from publicly available information.

Privacy of Relationships A user may have relationships with multiple users in the system. These other users may include fellow users who have rated the same ratees. The relationships between the users could be in different contexts, for example, the context of trust in preserving each others privacy. This requirement implies that information about the relationships of a rater is not revealed during the course of a reputation query. This information includes the amount of trust that the rater has in fellow users (other than the ratee), etc.

4.2.2 Correctness Objectives

No Out of Range Feedback A dishonest user is unable to submit out of range feedback. A dishonest user can take advantage of the fact that the feedback is confidential and submit out of range feedback in order to mount an attack such as bad mouthing or ballot stuffing.

No Incorrect Computations A dishonest user is unable to carry out incorrect computations. A reputation query may require users to perform certain computations, for example, the summation of some values. This property requires that a dishonest user is unable to submit erroneous results for these computations.

Observability of Adversarial Behavior Dishonest users that exhibit adversarial behavior (such as dropping messages, not participating in the query protocol, etc.) are revealed to honest users.

Termination The query protocol either terminates with the correct reputation score as the result or identifies the dishonest users that are hindering its termination.

4.3 Correctness Objectives Common to both types of Privacy Preserving Reputation Systems

Authorizability of Ratings The requirement of authorizability of ratings implies that only the users who have had a transaction with the ratee are allowed to rate him. This property prevents users who have not transacted with a ratee from assigning him feedback and thus possibly reduces the impact of attacks such as bad mouthing and self promotion.

Verifiability by Ratee The requirement of verifiability by ratee as identified by Kerschbaum [41] suggests that a ratee R should be able to identify all published feedback linked to his identity and verify that they are related to a recorded transaction and the correct transaction partners. Moreover, a ratee R should be able to identify all published feedback linked to his identity and verify that the inquirer has computed its reputation score according to them.

Verifiability by Rater Kerschbaum [41] states this property as follows: A rater R should be able to identify all published feedback linked to his identity and verify that the rating is as he left it.

5 Building Blocks for Privacy Preserving Reputation Systems

5.1 Secure Multi-Party Computation

Secure multi-party computation is the study of protocols that take inputs from distributed entities and aggregate them to produce outputs, while preserving the privacy of the inputs.

One of the well-known secure multi-party computation protocols is secure sum [17], which takes inputs from entities and computes their sum. The protocol may be used to compute reputation in the form of sum or mean while preserving the confidentiality of the feedback values under certain conditions.

5.1.1 Secure Sum

The protocol assumes that there are three or more sites and there is no collusion between them. It is also assumed that the value to be computed, $v = \sum_{l=1}^{s} v_l$ lies in the range $[0..m]$. The sites are numbered as $1 \ldots s$. Site

1 generates a random number R uniformly chosen from $[0..m]$. It then sends $R + v_1 \ mod \ m$ to site 2, where v_1 is site 1's local input. Site 2 does not learn any information about v_1 since $R + v_1 \ mod \ m$ is distributed uniformly across the range $[0..m]$ due to R. For sites $l = 2 \ldots s-1$, the protocol proceeds as follows: Site l receives:

$$V = R + \sum_{j=1}^{l-1} v_j \ mod \ m \tag{1}$$

Site l learns nothing since the value is distributed uniformly across $[0..m]$. Site l computes:

$$R + \sum_{j=1}^{l} v_j \ mod \ m = (v_l + V) \ mod \ m \tag{2}$$

Site l then sends this value to site $l+1$. Eventually, site s also performs the above step. Site s sends the result back to site 1, who subtracts R from it to obtain the sum. Site 1 does not learn any of the private values due to the uniform distribution of the received result over the range $[0..m]$.

The protocol may be used to compute reputation as the sum of the feedback values provided as private inputs by the participants of the protocol.

The security of the secure sum protocol does not hold if the sites are assumed to collude. Any two sites $l-1$ and $l+1$ can use the values that they send and receive respectively to compute the private input v_l of site l.

5.2 E-Cash

A number of privacy preserving reputation systems with user anonymity (such as [3]) use E-cash as one of the tools. E-cash is a digital currency first proposed by Chaum [15, 16]. E-cash provides the following features:

Anonymity It is impossible to trace an e-coin (the monetary unit of e-cash) to the user who spent it. This property holds even when the bank (a central entity who issues the e-coins) is the attacker.

Unforgeability The only exception to the anonymity property is that e-cash does not guarantee the anonymity of a user who tries to double-spend an e-coin. In this case, the bank can learn the identity of the dishonest user. A forged e-coin allows the bank to trace down the user who forged it.

Fungibility A user can use the e-coins received for services provided as payment for services received from any other user in the system.

Endorsed e-cash [13] adds the following property to e-cash:

Fair Exchange Fair exchange means that a buyer gets the item only if the seller gets paid and vice versa.

5.3 Trusted Third Party

A Trusted Third Party (TTP) for a set of agents is an entity whom every agent in the set fully trusts to preserve its privacy.

5.4 Anonymous Credential Systems

In anonymous credential systems (for example, [8, 12]), organizations grant credentials to pseudonymous identities of users. Verifiers are able to verify the authenticity of credentials in possession of users. However, neither an organization or a verifier is able to link a credential to the true identity of a user.

5.5 Blind Signatures

In a blind signature scheme (for example, [15]), an entity signs a message for a user, however the entity does not learn the content of the message.

5.6 Trusted Platform Modules

A Trusted Platform (TP) [47, 52] is described as a secure computing platform that preserves the privacy of the user by providing the following three functionalities:

Protected Storage Data on the TP is protected from unauthorized access.

Integrity The TP can prove that it is running only the authorized software and no malicious code.

Anonymity The TP can demonstrate that it is a genuine TP without revealing the identity of the user. The TP uses a pseudonym attested by a PKI Certification Authority (CA).

A Trusted Platform comprises of a Trusted Platform Module (TPM), which is a hardware device with cryptographic functions that enable the various security functionalities of the TP. The TPM is unforgeable and tamper-resistant.

5.7 Zero Knowledge Proofs

A zero-knowledge proof [30] is an interactive proof that allows a prover to convince a verifier that a statement is true without revealing any information other than the fact that the statement is valid.

As an example, consider a prover who knows an RSA modulus $\mathfrak{n}$ and its two large prime factors p and q. A verifier knows only $\mathfrak{n}$. Factoring $\mathfrak{n}$ is considered intractable therefore the verifier cannot learn p and q. An interactive proof would be zero-knowledge if it allows the prover to convince the verifier that he knows the factors of $\mathfrak{n}$ without revealing any information about p and q.

5.8 Homomorphic Cryptosystems

Let $E_a(.)$ denote an encryption function with the public key PK_a of agent a in an asymmetric cryptosystem $\mathcal{C}$. The cryptosystem $\mathcal{C}$ is said to be additive homomorphic if we can compute $E_a(x+y)$, given only $E_a(x)$, $E_a(y)$, and PK_a. In other words, a cryptosystem is additive homomorphic if we can compute the encryption of the sum of two plaintexts, given only their ciphertexts and the encrypting public key. As an example, let's consider two integers, 3 and 4. A cryptosystem $\mathcal{C}$ is additive homomorphic if given only $E_a(3)$, $E_a(4)$, and PK_a, we are able to obtain $E_a(3+4) = E_a(7)$. The Paillier cryptosystem [50] is a well-known additive homomorphic cryptosystem. Similarly, a multiplicative homomorphic cryptosystem such as the ElGamal Cryptosystem [27] allows computation of the encryption of the product of two plaintexts from their ciphertexts and the encrypting public key.

5.9 Anonymous Communication Networks

Anonymous communication networks, e.g., a Mixnet [8] or an Onion Router [24,11].

6 Privacy Preserving Reputation Systems in the Literature

We discuss several systems in the literature that relate to privacy preserving reputation systems. We summarize the salient features of each work, as well as present our analysis. A comparison of the systems is presented in Tables 1 through 7.

6.1 Pavlov et al. - Decentralized Additive Reputation Systems

Pavlov et al. [51] propose several protocols for decentralized additive reputation systems. Two of their protocols are secure under the semi-honest and the malicious adversarial models respectively. The protocols draw their strength from witness (feedback provider) selection schemes, which guarantee the inclusion of a certain number of honest witnesses as participants. The security mechanisms used in the protocols include secure multi-party computation, secret sharing, and discrete log commitment.

6.1.1 Problem Setting

A querying agent consults a group of n witnesses to compute the reputation of a target agent, where $0 < n < N$, and $N > 1$ is the number of potential witnesses. $b < N$ is the number of dishonest agents in N.

6.1.2 Decentralized Additive Reputation Systems

A decentralized additive reputation system is described in the article as a reputation system that satisfies the following two requirements: 1) feedback collection, combination, and propagation are implemented in a decentralized way; 2) combination of feedbacks provided by agents is calculated in an additive manner. The Beta reputation system [38] is cited as an example. The eBay reputation system is additive, however, not decentralized.

6.1.3 Impossibility of Perfect Privacy

The paper argues that it is impossible to guarantee perfect privacy for an honest feedback provider in a decentralized additive reputation protocol. The argument is that a dishonest agent may deterministically create a set of n feedback providers, with $n - 1$ dishonest agents and the one honest agent under attack. Given the inputs of the $n - 1$ dishonest agents and

the output (the reputation score), the secret feedback of the honest agent is easily obtained.

Analysis: The impossibility argument does not apply to protocols in which an honest agent may choose not to contribute his feedback. The argument also does not apply to protocols in which the set of feedback providers cannot be created deterministically.

6.1.4 Witness Selection Scheme 1 (WSS-1)

A witness selection scheme for a reputation protocol is a process that results in the creation of a set of witnesses. The witnesses in the set contribute their feedback towards computing the reputation of the target agent.

The first scheme [51, Lemma 2] guarantees that if honest agents are uniformly distributed over N, then at least two honest witnesses will be selected with probability greater than $(1 - \frac{1}{n})(\frac{N-b-1}{N-1})$. The scheme is secure under the semi-honest adversarial model, in which all agents follow the protocol correctly.

According to our analysis, the complexity of the number of messages exchanged is linear in terms of the number of potential witnesses: $O(N)$. After each witness is selected, it is probabilistically decided whether to add more witnesses, therefore the count may run up to N. If each agent sends its successor the current set of witnesses, the total bandwidth utilized is $O(N^2)$.

Analysis: The complexity of the scheme is a function of the population size of the potential witnesses (N) instead of the witnesses who contribute their feedback (n). The scheme also has the potential of leaving out many honest witnesses from the reputation protocol. Moreover, the scheme works only if $b < n - 1$, because otherwise $n - 1$ dishonest witnesses can select themselves into the set if the first witness selected is dishonest. Even then the scheme might fail since the number of witnesses selected is probabilistic and it may be the case that the actual number of selected witnesses is less than n.

6.1.5 Witness Selection Scheme 2 (WSS-2)

The second scheme [51, Lemma 3] guarantees under the malicious adversarial model that if honest agents are uniformly distributed over N, then at least $n(\frac{N-b-n}{N})$ honest witnesses would be selected. A coin flipping scheme is utilized to grow the set of witnesses by selecting the next witness randomly from the available pool of witnesses. According to the paper, the scheme

requires $O(n^3)$ messages among the n selected witnesses.

Analysis: It is not clear if the scheme would work in case the querying agent is dishonest. If the querying agent is dishonest, it does not need to follow the protocol correctly. It can select a dishonest witness and then collectively cheat to continue selecting dishonest witnesses. After an honest victim is selected, the rest of the witnesses must be selected randomly. However, at that point the coalition of dishonest agents has already biased the set in their favor.

6.1.6 A Reputation Protocol based on WSS-1

In this reputation protocol, the set of source agents is created using the first witness selection scheme, which guarantees that at least two source agents are honest. Agent q chooses a random number as its secret. Each agent splits its secret into $n+1$ shares such that they all add up to the secret. Each agent keeps the $n+1^{th}$ share and sends its other n shares to the other n agents in the protocol such that each agent receives a unique share. Each agent then adds all shares received along with his $n+1^{th}$ share and sends it to the querying agent. The querying agent adds all sums received and subtracts the random number to obtain the reputation score.

The protocol guarantees the privacy of an honest source agent under the semi-honest model as long as all the other $n-1$ source agents do not collude. The probability that all other source agents will not collude is greater than $(1-\frac{1}{n})(\frac{N-b-1}{N-1})$. The number of messages exchanged is analyzed as $O(n^2)$. We estimate that the size of the messages exchanged is as follows: $O(n^2)$ IDs and $O(n^2)$ numbers.

Analysis: The complexity is claimed to be $O(n^2)$, however, we believe it to be $O(N)+O(n^2)$ due to the utilization of the witness selection scheme.

6.1.7 A Reputation Protocol based on WSS-2

This protocol uses the Pedersen verifiable secret sharing scheme [53] and a discrete log commitment method. The Pedersen scheme is resilient up to $n/2$ malicious agents. The set of source agents is created using the second witness selection scheme. It guarantees the presence of less than $n/2$ malicious agents, if $b < \frac{N}{2} - n$.

The protocol is secure under the malicious model as long as $b < \frac{N}{2} - n$. The number of messages exchanged is $O(n^3)$, due to the second witness selection scheme.

6.2 Gudes et al. - The Knots Reputation System

Gudes et al. [31] present several schemes that augment their Knots reputation system [28] with privacy preserving features. A defining characteristic of the Knots reputation model is the notion of subjective reputation. The reputation of a target member is computed by each querying member using a different set of feedback, thus the reputation is subjective for each querying member. The feedback that a querying member uses for computing reputation comes exclusively from the members in which he has a certain amount of pre-existing trust. An advantage of this approach is that the querying member has confidence in each of the feedback values that are used for computing reputation.

Analysis: The disadvantage is that the opinion of the members whom the querying agent does not know in not taken into account. The notion of subjective reputation tends to be non-conformant with the idea of reputation, which is generally considered to be the aggregate of feedback of the community at large. The concept of subjective reputation seems closer to trust propagation than reputation.

6.2.1 The Knots Model

The Knots model differentiates between two types of users in the system. The *experts* in the system are the users who provide services and the *members* are users who consume those services. The reputation system is concerned with computing the reputation of the experts through the feedback provided by the members. Members have trust relationships among themselves in the context of providing reliable feedback about the experts.

$TrustSet_x(A)$ is defined as the set of members whom member A trusts to provide feedback about expert x. $TM(A, B)$ represents the amount of direct trust that a member A has in another member B. $DTE(A, x)$ is defined as the amount of direct trust that a member A has in an expert x. The subjective reputation of an expert x by a member A is computed as follows:

$$TE(A, x) = \frac{\Sigma_{\forall B \in TrustSet_x(A)} DTE(B, x) \cdot TM(A, B)}{\Sigma_{\forall B \in TrustSet_x(A)} TM(A, B)} \tag{3}$$

In the privacy preserving version of the Knots model, the challenge is to compute $TE(A, x)$, such that the privacy of each $DTE(B, x)$ is maintained, where $B \in TrustSet_x(A)$. The three decentralized privacy preserving schemes presented in the paper compute $\rho(A, x)$ (the numerator of the

fraction in equation 3), such that A cannot learn any of the $DTE(B, x)$ values.

Analysis: The privacy goal does not include preserving the privacy of the trust between the members (the TM values). It is limited to preserving the privacy of the feedback about the experts (the DTE values).

6.2.2 Reputation Scheme 1

Each member $B \in TrustSet_x(A)$ receives $TM(A, B)$ from A and then computes $E_A(DTE(B, x) \cdot TM(A, B))$ and sends it to a Trusted Third Party (TTP), Z (where $E_A(.)$ is an encryption with the public key of member A). The TTP Z relays each message to A without revealing the source member. A decrypts the messages and obtains $\rho(A, x)$.

Since A does not know the source of a message, it cannot reverse a received value to reveal the private feedback. The messages are encrypted, therefore the TTP does not learn any information either. The scheme requires $O(n)$ messages to be exchanged, where n is the cardinality of $TrustSet_x(A)$.

Analysis: The scheme requires disclosure of the trust that A has in each member B. Moreover, there is heavy reliance on the TTP. If the TTP and A collude, then they can easily determine each $TM(B, x)$.

6.2.3 Reputation Scheme 2

Each member $B \in TrustSet_x(A)$ generates $E_A(DTE(B, x))$ and sends it to a TTP, Z. The TTP sends a randomly permuted vector of the messages to A, who decrypts the messages and obtains a vector (vector 1) of the DTE values. A then sends a vector of all values $TM(A, B)$ to Z, where $B \in TrustSet_x(A)$. Z permutes the vector (vector 2) according to the DTE vector (with respect to the order of the members). A and Z compute the scalar product of vectors 1 and 2 using a secure product protocol (such as [2]) to obtain $\rho(A, x)$.

Due to the random permutation generated by the TTP, A is unable to correlate the DTE values with individual members. The TTP does not learn any of the DTE values due to encryption. A key advantage of the scheme is that any member B does not learn $TM(A, B)$.

We analyze that the number of messages exchanged is $O(n)$, whereas the bandwidth utilized is $O(n^2)$ in terms of k-bit numbers transfered, where k is the security parameter (key length).

Analysis: The privacy of the $TM(A, B)$ values is still not fully preserved since they must be disclosed to the TTP.

6.2.4 Reputation Scheme 3

A executes the reputation protocol for the semi-honest model from Pavlov et al. [51] to obtain $\Sigma_{\forall B \in TrustSet_x(A)} DTE(B, x)$. A sends $TM'(A, B) = TM(A, B) + Q$ to each $B \in TrustSet_x(A)$, where Q is a random number. A executes the secure sum protocol [17] to obtain $\Sigma_{\forall B \in TrustSet_x(A)}(TM'(A, B) \cdot DTE(B, x))$. A calculates:

$$\begin{aligned} \rho(A, x) = \Sigma_{\forall B \in TrustSet_x(A)}(TM'(A, B) \cdot DTE(B, x)) \\ -(Q \cdot \Sigma_{\forall B \in TrustSet_x(A)} DTE(B, x)) \end{aligned} \quad (4)$$

This scheme has the advantage that the privacy of both the $DTE(B, x)$ values and the $TM(A, B)$ values is preserved without the presence of any TTPs. The protocol requires $O(n^2)$ messages due to the inclusion of the protocol from [51].

6.2.5 Proposals for the Malicious Adversarial Model

The work also includes some proposals for augmenting the schemes for the malicious adversarial model.

Analysis: The proposals are largely based on the assumption that a member who provides feedback (member B) would lack the motivation to act maliciously if it does not know the identity of the querying member (member A). However, this assumption does not take into account the case when an attacker may want to attack the system simply to disrupt it, for example, in a denial-of-service attack.

6.3 Androulaki et al. - A Reputation System for Anonymous Networks

Androulaki et al. [3] propose a reputation scheme for pseudonymous peer-to-peer systems in anonymous networks. Users in such systems interact only through disposable pseudonyms such that their true identity is not revealed. Reputation systems are particularly important for such environments since otherwise there is little incentive for good conduct. However, reputation systems are hard to implement for these environments. One of the reasons is that a user must keep his reputation even if he cycles through many pseudonyms. Moreover, the pseudonyms must be unlinkable to the user as

well as to each other even though they share the same reputation score. Another issue that arises in reputation systems for anonymous networks is that a user may lend his good reputation to less reputable users through anonymous pseudonyms.

The proposed system employs the following cryptographic building blocks: anonymous credential systems, e-cash, and blind signatures. Reputation is exchanged in the form of e-coins called *repcoins*. The higher the amount of repcoins received from other users, the higher is the reputation of the user.

Analysis: The system requires the presence of a *bank*, which is a centralized entity. Additionally, the system also requires that all communication take place over an anonymous network, such as Mixnet [14] or a network using Onion routing [22]. This requirement makes the solution inaccessible to applications in non-anonymous networks.

The security goals of reputation systems for anonymous networks are different than those of privacy preserving reputation systems. The reputation systems for anonymous networks aim to hide the identity of a user who interacts and assigns feedback to others. Whereas, in privacy preserving reputation systems, the goal is to hide the feedback value assigned but not the identity of the user who assigned it. The choice between the two kinds of reputation systems depends on the security objectives of the application.

6.3.1 Security Model

Some of the security requirements of the reputation system are as follows:

Unlinkability An adversary, controlling the bank and a number of corrupted users, is unable to link a pseudonym with the identity of its non-corrupted user any better than by making a random guess. Moreover, the adversary has no advantage in telling whether two pseudonyms belong to the same non-corrupted user or not.

No Over-Awarding A user who tries to double-award (forge) a repcoin, using one or even two different pseudonyms, gets detected and his identity is revealed.

Exculpability Any coalition of corrupted users (including the bank) is unable to falsely accuse a user of forgery in order to expose to his identity.

Reputation Unforgeability, Non-Transferability A user cannot forge better reputation. In particular, a user U_1 cannot borrow reputation from another user U_2, unless U_2 reveals his master secret key to U_1.

6.3.2 A Reputation System for Anonymous Networks

The system assumes the presence of a central entity called the bank, which is needed for implementing the above listed cryptographic schemes. The system also requires that all communication takes place over an anonymous network, for example, a Mixnet, or a network using Onion routing. The users interact with each other in a peer-to-peer manner. However, the users must also communicate with the central bank to withdraw and deposit repcoins.

From the above listed building blocks, Androulaki et al. build a reputation system in which each user has a reputation that he cannot lie about or shed. However, a user may generate as many one time pseudonyms as he needs for his transactions. All pseudonyms of a user share the same reputation. The system is robust against self-promotion attacks. Reputation is updated and demonstrated in a way such that anonymity is not compromised. The system maintains unlinkability between the identity of a user and his pseudonyms, and unlinkability among pseudonyms of the same user.

The system by Androulaki et al. follows upon the work by Dingledine et al. [24, 23, 21] on reputations systems and anonymous networks.

6.4 Nin et al. - A Reputation System for Private Collaborative Networks

Nin et al. [48] present a reputation system that computes the reputation of a user based on the access control decisions that he makes. If a user makes good access control decisions, such as granting access to legitimate users and denying access to unauthorized users, then he receives good reputation. In contrast, making dishonest access control decisions leads to bad reputation. The privacy objective of the reputation system is to keep the trust relationships between the users private.

The system operates as follows: A node keeps record of its access control decisions. Other nodes can view anonymized details of those decisions and verify if the decisions were made according to the access control rules or not. The anonymization is derived through the multiplicative homomorphic property of the ElGamal encryption scheme. Private details are not revealed to a third-party due to the anonymization.

6.4.1 Private Collaborative Networks

A private collaborative network is described as a network of users that has the following properties: 1) the users are connected with each other through trust relationships; 2) users own resources that can be accessed by other

users if sufficient trust exists; and 3) trust relationships among users remain private.

A private collaborative network is modeled as a directed labeled graph. Edges represent trust relationships between nodes (users). Each edge is labeled with the type of trust relationship as well as the weight of the trust.

Access to each resource in the network is governed by a set of access conditions. An access condition is of the form $ac = (v, rt, d_{max}, t_{min})$, where v is the owner with whom the requester of the resource must have a direct or transitive trust relationship of type rt to gain access. d_{max} and t_{min} are the required maximum depth and minimum trust respectively to obtain access.

Each trust relationship also exists in the form of a certificate signed by the truster and the trustee. Since relationships must be kept private, a certificate itself is considered a private resource. To gain access to a resource, a requester must demonstrate to the owner, the existence of a "certificate path" linking the requester to the owner.

6.4.2 The Reputation Model

The reputation system assigns good reputation to a user who performs decisions in accordance with the specified access conditions. In contrast, a user who does not correctly enforce access control rules, receives lower reputation. Reputation lies in the interval $[0, 1]$.

A user can act dishonestly in two ways: 1) deny access to a resource to a legitimate requester, or 2) allow access to a resource to an unauthorized requester. The access control decision is considered wrong if it violates either of the rt, d_{max}, t_{min} parameters in the access condition. For a wrong decision that violates the trust requirement (t_{min}), the absolute difference between the minimum amount of trust required (t_{min}) and the trust computed over the certificate path is given as wd. The values arising from all such wrong decisions are given as the set $\{wd_1, \ldots, wd_{|WD_{t_A}|}\}$, where $|WD_{t_A}|$ is the number of wrong decisions.

The values in the set $\{wd_1, \ldots, wd_{|WD_{t_A}|}\}$, which represent the wrong decisions made by user A in terms of trust, are aggregated as:

$$AGt_{AC_{SET_A}} = OWA_Q(wd_1, \ldots, wd_{|WD_{t_A}|}) \tag{5}$$

where $AGt_{AC_{SET_A}}$ is the aggregated value of the wrong decisions with respect to trust. OWA is an Ordered Weighted Averaging function and Q is a non-decreasing fuzzy quantifier. According to the authors: "The interest of the OWA operators is that they permit the user to aggregate the values giving importance to large (or small) values".

The wrong decisions of the user that violate the depth and path requirements are aggregated as $AGd_{AC_{SET_A}}$ and $AGp_{AC_{SET_A}}$ respectively. The reputation of user A is then computed as:

$$R_A = 1 - \frac{1}{3}(AGt_{AC_{SET_A}} + AGd_{AC_{SET_A}} + AGp_{AC_{SET_A}}) \tag{6}$$

which implies that the mean of the aggregates of the three types of wrong decisions is subtracted from the perfect reputation of 1 to arrive at the reduced reputation of the user. The more dishonest decisions a user makes, the lower his reputation.

6.4.3 Anonymized Audit Files

After a user makes an access control decision, an entry about that decision is added into the user's anonymized audit file. The entry includes information such as the identity of the requester of the resource, the certificate path demonstrated by the requester, etc. However, all private information in the entry is encrypted using the ElGamal encryption scheme [27]. Therefore, a third-party who analyzes the entry is unable to acquire any information about these private elements. Due to the multiplicative homomorphic nature of the ElGamal encryption scheme, the encrypted information can be manipulated to compute reputation. A network participant who wishes to learn the reputation of a certain user, can analyze the anonymized audit file of that user and derive the reputation score without compromising privacy.

We analyze the number of messages exchanged to compute reputation as constant ($O(1)$), since all required information is provided directly by the target node.

Analysis: We believe that the following features of the reputation system are advantageous: 1) the reputation of a node is not derived from the feedback of other nodes but from objective information about its behavior (its access control decisions), and 2) a node itself manages and furnishes the evidence required for another node to judge its reputation.

However, we also observe the following issues: 1) As we understand, a node itself manages its audit file due to the absence of centralized entities and TTPs in the system. It is not clear why a dishonest node would include its bad decisions in its audit file. If the node is itself in charge of creating the file, it would only include details that lead to good reputation. 2) Reputation is computed based on the access control decisions of a user, which makes the applicability of the reputation system limited. For example, the reputation

system would not work in e-commerce systems, where the reputation of a seller is based on the subjective satisfaction of the buyers.

The adversarial model is not specified in the paper, however, we estimate that the scheme would be secure only upto the semi-honest model since nodes are assumed to manage their audit files honestly.

6.5 Kinateder and Pearson - A Privacy-Enhanced P2P Reputation System

The decentralized reputation system proposed by Kinateder and Pearson [42] requires a Trusted Platform Module (TPM) chip at each agent. The TPM enables an agent to demonstrate that it is a valid agent and a legitimate member of the reputation system without disclosing its true identity. This permits the agent to provide feedback anonymously.

6.5.1 Security Goals

The reputation system sets the security requirements listed below. An attacker must not be able to:

- Provide false feedback on an honest user's behalf.
- Access an honest user's private database and modify data such as feedback, reputation, etc.
- Learn the identity of a feedback provider (which implies that a user should be able to provide feedback anonymously).

Moreover, it is required that:

- The identity of a dishonest user can be revealed if there is sufficient legal justification.

6.5.2 System Model and Functionality

An agent in the system can take up one of following three roles at any given time: *recommender*, *requester*, and *accumulator*.

Recommender A recommender agent has interacted directly with other agents and has feedback about them. He regularly announces the availability of feedback to other agents in the system. A recommendation comprises of the target agent's pseudonym, the recommender agent's

pseudonym, and the feedback value. The recommendation is digitally signed by the recommender.

Accumulator An accumulator agent stores feedback about other agents. However, his feedback is not based on direct experience with the target agent but formed through the feedback that he has received from other agents in the system.

Requester A requester agent queries other agents for feedback and then locally aggregates the feedback to determine the reputation of the target agent. A requester agent propagates the query to its peer agents who in turn propagate to their peer agents. Each peer decides when to discontinue further propagation based on whether recommendations are available among its peers. The requester agent receives the feedback from the recommender and accumulator agents queried and then aggregates the feedback to learn the reputation of the target agent.

Analysis: It is not elaborated how the feedback announcement and feedback query protocols work, for example, if an algorithm such as broadcast or gossip is used. As a consequence, the complexity of the protocols is not clear. Moreover, the mechanism for aggregating the feedback is not discussed.

6.5.3 How Security is Achieved

The security requirements are fulfilled as follows:

- An attacker is unable to provide false feedback on an honest user's behalf since each feedback is digitally signed by the recommender. A requester agent can also verify through the recommender's TP that it has not been compromised by the adversary.
- An attacker is unable to access an honest user's private database and modify data such as feedback, reputation, etc. This is achieved due to the protected data storage functionality of the TP. Therefore, a requester can be certain that the given feedback is not false.
- An attacker does not learn the true identity of a feedback provider since only pseudonyms are used. Thus, a user is able to provide feedback anonymously and without inhibition. The pseudonym is protected by the TP and the CA of the user. Moreover, the use of MIX cascades is suggested to prevent the attacker from correlating the pseudonym with the IP address of the user.

- In case of legal justification, the CA of a user can reveal his true identity.

Voss et al. [61] and Bo et al. [9] also present decentralized systems that are based on similar lines. They both suggest using smart cards as the trusted hardware modules. A later system by Kinateder et al. [43] avoids the hardware modules, however, it requires an anonymous routing infrastructure at the network level.

Analysis: Consider the following scenario: A sale on an e-commerce system may result in the disclosure of the true identities of the seller and the buyer to each other (through mailing addresses etc.), even if they use anonymous pseudonyms. We must also consider that the privacy of the pseudonym itself may need to be protected. For example, if pseudonym A assigns pseudonym B negative feedback in retaliation, then B's reputation is adversely affected due to the lack of privacy of B's feedback. Better solutions include: preserving the privacy of the feedback, or using disposable pseudonyms, which a user may change after every transaction (such as in the solution by Androulaki et al. [3]).

6.6 Steinbrecher - Privacy-Respecting Reputation Systems within Centralized Internet Communities

Steinbrecher [58] argues that traditional cryptographic techniques such as encryption and digital signatures can provide only "technical" security guarantees. For example, encryption and digital signatures can guarantee the confidentiality and integrity of the *text* of a reply sent by an expert to a user on a self help forum. However, these techniques cannot guarantee the misbehavior of the users themselves. For example, the user might violate confidentiality by relaying the *content* of the text to a third party, or the expert may violate integrity by giving *false advice*. It is argued that trust can mitigate these risks and that reputation systems are a suitable technology for acquiring trust.

However, the author contests that the design of current reputation systems (such as the eBay reputation system) allow open access to the interests and behavior profiles of users. A third-party may acquire information such as the time and frequency of participation, interests in specific items, feedback provided etc. Moreover, it is easy to associate the pseudonym of a user with their real identity, for example, through a mailing address.

To counter this issue, Steinbrecher presents a privacy-respecting reputation system for centralized Internet communities. The system relies on

simultaneous use of multiple pseudonyms and changing them frequently to achieve anonymity and unlinkability.

6.6.1 A Generalized Model for Centralized Reputation Systems

The paper presents a generalized model for centralized reputation systems. Users use global pseudonyms tied to global reputations. The set of global pseudonyms at time t is considered as $P_t = \{p_{t,1}, \ldots, p_{t,m}\}$. The set of possible reputations that might be associated with a pseudonym is given as R. $(R, +)$ is a commutative group and $+$ an operator to combine elements from R independently of t. At time t_1, each pseudonym $p_{t_1,l}$ has the reputation $rep(t_1, p_{t_1,l}) \in R$, where $l \in 1 \ldots m$. After $p_{t_1,i}$ receives a rating r_{j,i,t_1} from $p_{t_1,j}$, the reputation of $p_{t_1,i}$ at time t_2 is computed as:

$$rep(t_2, p_{t_1,i}) = rep(t_1, p_{t_1,i}) + r_{j,i,t_1} \tag{7}$$

where $t_2 \geq t_1$, and $p_{t_1,i}$ does not receive any rating other than r_{j,i,t_1} between t_1 and t_2.

6.6.2 Using Pseudonyms for Unlinkability and Anonymity

The system proposes simultaneous use of multiple pseudonyms by a user. The idea is to have a separate pseudonym for each context (for example, the context of a seller on an auction site, the context of an expert on a self help forum, etc.). It is suggested that this design leads to unlinkability between the different roles of a user on the Internet.

The system permits users to regularly change their pseudonyms to achieve anonymity. A new and an old pseudonym are unlinkable from the perspective of third-parties, however, the provider (central server) is able to link the two pseudonyms. The unlinkability also assumes that a large number of pseudonyms have the same reputation.

To prevent the provider from linking new and old pseudonyms, the system suggests using a set of non-colluding trustworthy third parties who make incremental changes to the pseudonym of the user.

Steinbrecher's work on reputation and privacy also includes [56, 54]. These proposals are oriented for centralized environments as well.

Analysis: An adversary may compromise unlinkability by monitoring all pseudonyms with the same reputation. The adversary can deduce that a new pseudonym with the same reputation as a recently deleted pseudonym belong to the same user.

6.7 Hasan et al. - The k-Shares Reputation Protocol

The k-shares protocol offers the following advantages over comparable protocols such as those by Pavlov et al. [51, Section 5.2] and Gudes et al. [31]: 1) Lower message complexity of $O(n)$ as opposed to $O(n^2)$ and above of the protocols in [51] and [31]; 2) The k-Shares protocol allows agents to quantify and maximize the probability that their privacy will be preserved before they submit their feedback.

6.7.1 Framework

The environment is modeled as a multi-agent environment. Subscribing to the definition of trust by sociologist Diego Gambetta [29], trust is characterized as binary-relational, directional, contextual, and quantifiable as subjective probability. Thus the feedback that a source agent assigns a target agent is considered as trust in a certain context and is quantified as probability that is subjective to the source agent. The context of the trust is an action and the quantification is the subjective probability that the target agent will perform that action. Reputation of a target agent is defined as any function that aggregates the feedback of its source agents. The protocol realizes reputation with the mean function, which is derived from summation.

A special action called "preserve privacy" is defined. Agents are assumed to have trust relationships with a some other agents in the context of this action. This assumption derives from the fact that agents have social relationships and a key component of such relationships is the trust that each others privacy will be preserved. For example, a user may trust its family members and close friends to help him preserve his privacy.

The adversary is considered as semi-honest and is allowed to collude. The paper also proposes ideas for adapting the protocol to the malicious adversarial model as part of future work. Privacy is formalized using the Ideal-Real approach. An ideal protocol for computing reputation is one in which a Trusted Third Party (TTP) receives all inputs and then locally computes the reputation. On the other hand, a real protocol computes reputation without the participation of any TTP. The real protocol is said to preserve privacy if the adversary, with high probability, cannot obtain any more information about the private input of an agent than it can learn in the ideal protocol.

6.7.2 The Protocol

A simplified version of the protocol is outlined below.

1. **Initiate.** The querying agent q retrieves the set of source agents S_t of the target agent t and sends the set to each of the source agents.

2. **Select Trustworthy Agents.** Each source agent selects up to k other agents in S_t. Each agent selects these agents such that the probability that all of them will collude to break his privacy is low. k is a constant, such that $k \ll n$, where n is the number of all source agents. The risk to privacy is thus quantified before submitting the feedback.

3. **Prepare and Send Shares.** Each agent generates k shares such that their sum is equal to the secret feedback value. The secret cannot be revealed until all shares are known. The shares are sent to the selected fellow agents.

4. **Compute Sums and Reputation.** Each agent that receives shares from fellow agents computes the sum of all shares received and sends the sum to the querying agent q. Agent q receives all the sums and computes the grand total and divides it by n to learn the reputation score.

The full version of the protocol takes measures to ensure that a share is not compromised even if it is the only share received by an agent. Moreover, the protocol also takes steps so that the protocol does not reach certain failure states.

The highlights of the protocol are as follows: 1) It requires each source agent to send only $k \ll n$ messages, which implies that the protocol requires only $O(n)$ messages. 2) The risk to privacy can be quantified before submitting feedback. Thus, an agent knows the risk and if that risk is unacceptable it can opt to not participate in the protocol. As a consequence, even up to $n-1$ dishonest agents in the protocol cannot breach the privacy of one dishonest agent.

6.7.3 Experimental Results

The paper conducts experiments on the real web of trust of Advogato.org. The members of Advogato rate each other in the context of being active and responsible members of the open source software developer community. The choice of feedback values are *master*, *journeyer*, *apprentice*, and *observer*,

with *master* being the highest level in that order. The result of these ratings is a rich web of trust. The members of Advogato are expected to not post spam, not attack the Advogato trust metric, etc. It is therefore argued that the context "be a responsible member of the open source software developer community" comprises of the context "be honest". The four feedback values of Advogato are substituted as follows: $master = 0.99$, $journeyer = 0.70$, $apprentice = 0.40$, and $observer = 0.10$. For the experiments, the lowest acceptable probability that privacy will be preserved is defined as 0.90. This means that a set of two trustworthy agents must include either one *master* rated agent or two *journeyer* rated agents for this security threshold to be satisfied. The two experiments in the paper and their results are as follows:

Experiment 1: In the k-Shares protocol, the following assumption must hold for an agent $\boldsymbol{a}$'s privacy to be preserved: the probability that the agents to whom agent $\boldsymbol{a}$ sends shares, are all dishonest must be low. The experiment determines the percentage of instances of source agents in the Advogato data set for whom this assumption holds true.

Results: Consider the case where there are at least 50 source agents present in the protocol and $k = 2$, that is only two trustworthy agent can be selected to preserve privacy. It is observed that the assumption holds for 85.8% of instances of source agents. At $n \geq 5$, the percentage is 72.5%.

Experiment 2: The experiment observes the effect of increasing k on the percentage of instances of source agents whose privacy is preserved by the k-Shares protocol in the Advogato.org data set.

Results: Consider the case where there are at least 50 source agents present in the protocol and $k = 1$, that is only one trustworthy agent can be selected to preserve privacy. In the percentage of instances of source agents whose privacy is preserved is 75.4%. At $k = 2$, the percentage is 85.8%. The rise is due to the possibility with $k = 2$ to rely on two trustworthy agents. Increasing k over 2, even up to 500, does not result in a significant advantage (86.3% at $k = 500$). These results validate the assumption that the privacy of a large number of agents can be preserved with $k \ll n$.

6.8 Comparison

Tables 1 through 7.

Table 1: Analysis of Reputation Systems - Fundamentals.

System	**Architecture**	**Feedback**		**Reputation**					**Feedback Aggregation Model**
		Set / Range	Dynamicity	Set / Range	Liveliness	Visibility	Durability	Monotonicity	
eBay (`ebay.com`)	C	$\{-1, 0, 1\}$	N	$\mathbb{Z}$	Y	G	Y	N	Sum
EigenTrust [40]	D	$[0, 1]$	Y	$\mathbb{R}$	Y	L	N	N	Markov Chain, Left principal eigenvector
Advogato [44]	C	{Apprentice, Journeyer, Master}	Y	{Apprentice, Journeyer, Master}	Y	G	Y	N	Flow Network
Beta [38]	C	$\{-1, 1\}$	N	$\mathbb{Z}$	Y	G	Y	N	Bayesian
Pavlov et al. [51]	D	$\mathbb{R}$	Y	$\mathbb{R}$	Y	L	N	N	Sum
Gudes et al. [31]	D	$\mathbb{R}$	Y	$\mathbb{R}$	Y	G	N	N	Sum
Kinateder et al. [42]	D	$\{-1, 0, 1\}$	Y	$\mathbb{Z}$	Y	L	Y	N	Sum
Androulaki et al. [3]	H	$\{0, 1\}$	N	$\mathbb{N}$	Y	G	Y	Y	Sum
Nin et al. [48]	D	$\{-1, 0, 1\}$	N	$\mathbb{Z}$	Y	G	Y	N	Sum
Steinbrecher [58]	C	$\{-1, 0, 1\}$	N	$\mathbb{Z}$	Y	G	Y	N	Sum
Hasan et al. [35]	D	$[0, 1]$	Y	$\mathbb{R}$	Y	G	N	N	Sum, Mean

Table 2: Legend for Table 1.

Column	**Symbol**	**Description**
Architecture	C	Centralized
	D	Decentralized
	H	Hybrid
Feedback, Reputation	Y	Yes
	N	No
Reputation // Visibility	G	Global
	L	Local

Table 3: Analysis of Reputation Systems - Measures Against Challenges - Y: Yes, N: No, P: Partial.

System	Privacy	Sybil Attack	Ballot Stuffing	Slandering	Whitewashing	Oscillation
eBay (`ebay.com`)	N	P	P	Y	P	N
EigenTrust [40]	N	Y	Y	P	Y	P
Advogato [44]	N	Y	Y	P	Y	P
Beta [38]	N	N	Y	Y	N	N
Pavlov et al. [51]	Y	N	N	N	N	N
Gudes et al. [31]	Y	N	N	N	N	N
Kinateder et al. [42]	Y	P	P	N	P	N
Androulaki et al. [3]	Y	N	N	N	P	N
Nin et al. [48]	Y	N	N	N	N	N
Steinbrecher [58]	Y	P	P	N	P	N
Hasan et al. [35]	Y	P	N	N	N	N

Table 4: Analysis of Reputation Systems - Reputation Query Costs - n: No. of raters, N: No. of all users.

System	Messages
eBay (`ebay.com`)	$O(n)$, $\Omega(1)$
EigenTrust [40]	$O(logN)$
Advogato [44]	$O(n)$, $\Omega(1)$
Beta [38]	$O(n)$, $\Omega(1)$
Pavlov et al. [51]	$O(n^2) + O(N)$
Gudes et al. [31]	$O(n^2) + O(N)$
Kinateder et al. [42]	-
Androulaki et al. [3]	$O(n)$, $\Omega(1)$
Nin et al. [48]	$O(n)$, $\Omega(1)$
Steinbrecher [58]	$O(n)$, $\Omega(1)$
Hasan et al. [35]	$O(n)$

7 Discussion

The tables provide a comparison of the reputation systems that aim to preserve privacy under the semi-honest adversarial model and the disruptive

Table 5: Analysis of Privacy Preserving Reputation Systems - Fundamentals - SH: Semi-Honest, M: Malicious, Y: Yes, N: No, P: Pseudonym, I: Identity, UA: User Anonymity, FC: Feedback Confidentiality.

System	**Adversarial Model**	**Collusion**	**Reputation Binding**	**Privacy Objective**	**Building Blocks**
Pavlov et al. [51]	S	Y	P	FC	SMPC
Gudes et al. [31]	S	Y	P	FC	SMPC
Kinateder et al. [42]	M	Y	I	UA	TPM
Androulaki et al. [3]	M	Y	I	UA	E-Cash
Nin et al. [48]	S	Y	P	FC	Homomorphic crypto.
Steinbrecher [58]	M	Y	I	UA	ID management
Hasan et al. [35]	S	Y	P	FC	SMPC, Trust

Table 6: Analysis of Privacy Preserving Reputation Systems - Privacy Objectives - User Anonymity Oriented Systems.

System	**Privacy**							**Integrity**			
	Multiple Pseudonyms	User-Pseudo Unlinkability	Pseudo-Pseudo Unlinkability	Rater Anonymity	Ratee Anonymity	Inquirer Anonymity	Reputation Transfer	Unforeability	Distinctness	Verifiability	Authorizability
Kinateder et al. [42]	N	Y	-	Y	Y	N	-		N	N	N
Androulaki et al. [3]	Y	Y	Y	Y	Y	N	Y	Y	N	N	N
Steinbrecher [58]	Y	Y	P	Y	Y	N	Y	Y	N	N	N

Table 7: Analysis of Privacy Preserving Reputation Systems - Privacy Objectives - Feedback Confidentiality Oriented Systems.

System	Privacy				Integrity			
	Feedback Confid. (Inter.)	Feedback Confid. (Public.)	Trust Relationships	Ideal vs Real	Correct Range	Correct Computation	Verifiability	Authorizability
Pavlov et al. [51]	Y	N	Y	Y	Y	Y	N	N
Gudes et al. [31]	Y	N	N	Y	Y	Y	N	Y
Nin et al. [48]	Y	N	N	N	Y	Y	N	Y
Hasan et al. [35]	Y	N	P	Y	Y	Y	N	Y

malicious adversarial model respectively.

7.1 The Semi-Honest Adversarial Model

The Secure Sum protocol is simple and efficient. However, secure sum is secure only under a restricted semi-honest adversarial model where the entities are not allowed to collude. The protocol is therefore not suitable for preserving privacy under the more realistic model where collusion is possible.

The schemes 1 and 2 by Gudes et al. provide security under the full semi-honest model. However, both schemes rely on Trusted Third Parties (TTPs). The issue with TTPs is that if they are not fully honest, they can learn private data with little or no effort.

The reputation system by Nin et al. is very efficient. It requires exchange of a constant number of messages. However, the system is limited to Private Collaborative Networks, where reputation is computed based on the access control decisions of an entity. The reputation system is not applicable to more general areas, such as e-commerce, peer-to-peer file sharing, etc.

The protocol by Pavlov et al. (based on their first witness selection scheme) is secure under the full semi-honest model. Moreover, the protocol is general purpose, that is, it may be used for many different applications. The protocol also does not rely on any TTPs or centralized constructs. The

scheme 3 by Gudes et al. has similar properties. However, both these protocols have communication complexity upwards of $O(n^2)$, which is quite expensive.

7.2 The Disruptive Malicious Adversarial Model

The reputation systems by Androulaki et al. and Steinbrecher are very efficient. They require a constant number of messages to be exchanged despite the number of feedback providers and the size of the system. However, each of these systems relies on a centralized construct. The reputation system by Androulaki et al. is based on the E-Cash system, which uses a centralized construct called the bank. Steinbrecher's reputation system has a central server as an integral part of its architecture. These centralized entities make these two systems unsuitable for fully decentralized environments.

Kinateder et al.'s reputation system provides anonymity in peer-to-peer systems under the disruptive malicious model. However, the system requires the presence of special hardware called Trusted Platform (TP) at each peer. Additionally, the system requires that messages be exchanged using MIX cascades. These requirements limit the reputation system to specialized networks where TPs are available at each peer and where MIX cascades are in use.

The protocol by Pavlov et al. (based on their second witness selection scheme) is secure under the disruptive malicious model. The protocol does not require centralized constructs or specialized networks. However, the issue with the protocol is that it needs $O(n^3)$ messages to be exchanged, which is very expensive.

8 Related Work

Schiffner et al. [56] present an analysis of some privacy preserving reputation systems in the literature as part of their paper that describes a novel system that preserves privacy as well maintains liveliness. Their analysis compares their own system with two other systems, namely the systems by Androulaki et al. [3] and Voss [61]. In contrast, our survey presents an analysis framework that covers a wide array of privacy preserving reputation systems. Moreover, we analyze and compare different privacy preserving reputation systems belonging to the two different categories of user anonymity and feedback confidentiality.

Hoffman et al. [36] present a survey of attack and defense techniques for reputation systems. The survey describes a number of challenges that

reputation systems face and techniques that can resolve those challenges. However, their work does not address the issue of privacy in reputation systems. Another survey by Marmol and Perez [45] also analyzes threat scenarios for reputation systems. However, their survey also does not cover privacy preserving reputation systems.

9 Conclusion

This survey of privacy preserving reputation systems makes the following contributions:

- Identification of the various dimensions of privacy preserving reputation systems. An analysis framework that allows for the decomposition and comparison of privacy preserving reputation systems in a normalized manner. As a first step, we presented an analysis framework that covers the fundamental elements of reputation systems that are common to all reputation systems and not just those that preserve privacy. We identified the following elements for this initial framework: the architecture of the system, the nature of the feedback, the nature of the reputation, the feedback aggregation model, the challenges addressed, and the reputation query costs. The subsequent analysis framework that specifically addresses privacy preserving reputation systems decomposes them according to the following dimensions: the nature of the adversary, reputation binding, the privacy objective of the system, and the building blocks utilized.

- Identification of the privacy requirements of privacy preserving reputation systems that cut across multiple types of such systems. We identified that there are two main types of privacy preserving reputation systems: 1) systems that preserve the anonymity of the users, and 2) systems that don't preserve the anonymity of the users but preserve the confidentiality of their feedback. We further identified that the privacy-related requirements can be further subdivided into privacy requirements and integrity requirements. The requirements that we have identified as part of a unified framework can serve as a guide for designers of privacy preserving reputation systems.

- Identification of the building blocks of current privacy preserving reputation systems. Identification of the privacy preserving strategies using the building blocks and their respective advantages and disadvantages. We observed that the various strategies and associated building

blocks offer individual advantages and disadvantages. For example, the E-Cash strategy can be used to preserve the anonymity of users, however, it requires a centralized entity which makes it unsuitable for decentralized networks. A designer of a privacy preserving reputation system can use this analysis to select the best-suited building blocks.

- Comparison of privacy preserving reputation systems as well as prominent non-privacy preserving reputation systems using the analysis framework. Our detailed comparison of privacy preserving reputation systems in a normalized manner systems using our analysis framework reveals the differences between the systems in the literature. Another important contribution is the comparison of privacy preserving and non-privacy preserving systems, which reveals the gap between these two systems.

- Review of representative privacy preserving reputation systems in the literature. Identification of individual strengths and weaknesses.

- Analysis of the effect of preserving privacy on other challenges faced by reputation systems. Identification and discussion of open issues in privacy preserving reputation systems. We observe that preserving privacy raises issues in addressing other challenges such as slandering or bad-mouthing. We also observed that most current systems use the primitive sum function as feedback aggregation model. Developing privacy preserving systems that utilize other aggregation functions could lead to systems that address multiple challenges.

References

[1] Aladdin Knowledge Systems Ltd. Attack intelligence research center annual threat report – 2008 overview and 2009 predictions. http://www.aladdin.com/pdf/airc/AIRC-Annual-Threat-Report2008.pdf, 2008.

[2] Artak Amirbekyan and Vladimir Estivill-Castro. A new efficient privacy-preserving scalar product protocol. In *Proceedings of the Sixth Australasian Conference on Data Mining and Analytics*, 2007.

[3] Elli Androulaki, Seung Geol Choi, Steven M. Bellovin, and Tal Malkin. Reputation systems for anonymous networks. In *Proceedings of the 8th Privacy Enhancing Technologies Symposium (PETS 2008)*, 2008.

[4] Mohd Anwar and Jim Greer. Reputation management in privacy-enhanced e-learning. In *Proceedings of the 3rd Annual Scientific Conference of the LORNET Research Network (I2LOR-06)*, Montreal, Canada, November 2006.

[5] Mohd Anwar and Jim Greer. Enabling reputation-based trust in privacy-enhanced learning systems. In *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, Montreal, Canada, 2008.

[6] Roberto Aringhieri, Ernesto Damiani, Sabine De Capitani Di Vimercati, Stefano Paraboschi, and Pierangelo Samarati. Fuzzy techniques for trust and reputation management in anonymous peer-to-peer systems. *Journal of the American Society for Information Science and Technology*, 57(4):528537, February 2006.

[7] Mira Belenkiy, Melissa Chase, C. Chris Erway, John Jannotti, Alptekin Kupcu, Anna Lysyanskaya, and Eric Rachlin. Making p2p accountable without losing privacy. In *Proceedings of the 2007 ACM Workshop on Privacy in Electronic Society*, 2007.

[8] Mira Belenkiy, Melissa Chase, Markulf Kohlweiss, and Anna Lysyanskaya. P-signatures and noninteractive anonymous credentials. In *Theory of Cryptography*, 2008.

[9] Yang Bo, Zhou Min, and Li Guohuan. A reputation system with privacy and incentive. In *Proceedings of the Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD'07)*, 2007.

[10] Sonja Buchegger and Jean-Yves Le Boudec. Performance analysis of the confidant protocol (cooperation of nodes: Fairness in dynamic ad-hoc networks). In *Proceedings of the Third ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC'02)*, Lausanne, Switzerland, June 2002.

[11] Sonja Buchegger and Jean-Yves Le Boudec. A robust reputation system for peer-to-peer and mobile ad-hoc networks. In *Proceedings of P2PEcon 2004*, Harvard University, Cambridge, MA, USA, June 2004.

[12] J. Camenisch and A. Lysyanskaya. An efficient system for non-transferable anonymous credentials with optional anonymity revocation. In *EUROCRYPT 2001*, 2001.

[13] J. Camenisch, A. Lysyanskaya, and M. Meyerovich. Endorsed e-cash. In *Proceedings of the IEEE Symposium on Security and Privacy*, 2007.

[14] D. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 24(2):8488, 1981.

[15] D. Chaum. Blind signatures for untraceable payments. In *Proc. Advances in Cryptology (CRYPTO '82)*, 1982.

[16] D. Chaum. Blind signature systems. In *Advances in Cryptology (CRYPTO'83)*, 1983.

[17] Chris Clifton, Murat Kantarcioglu, Jaideep Vaidya, Xiaodong Lin, and Michael Y. Zhu. Tools for privacy preserving distributed data mining. *SIGKDD Explorations*, 4(2):28–34, January 2003.

[18] C. Costa and J. Almeida. Reputation systems for fighting pollution in peer-to-peer file sharing systems. In *Proceedings of the Seventh IEEE International Conference on Peer-to-Peer Computing*, October 2007.

[19] CyberSource Corporation. Cybersource 12th annual online fraud report. http://www.cybersource.com/, 2011.

[20] Ernesto Damiani, Sabrina De Capitani di Vimercati, Stefano Paraboschi, Pierangela Samarati, and Fabio Violante. A reputation-based approach for choosing reliable resources in peer-to-peer networks. In *Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS'02)*, 2002.

[21] R. Dingledine, M. J. Freedman, D. Hopwood, and D. Molnar. A reputation system to increase mix-net reliability. In *Proceedings of the 4th International Workshop on Information Hiding*, 2001.

[22] R. Dingledine, N. Mathewson, and P. F. Syverson. Tor: The second-generation onion router. In *Proceedings of the USENIX Security Symposium*, 2004.

[23] Roger Dingledine, Nick Mathewson, and Paul Syverson. Reputation in privacy enhancing technologies. In *Proceedings of the 12th Annual Conference on Computers, Freedom and Privacy*, 2002.

[24] Roger Dingledine, Nick Mathewson, and Paul Syverson. Reputation in p2p anonymity systems. In *Proceedings of the Workshop on Economics of Peer-to-Peer Systems*, 2003.

[25] John R. Douceur. The sybil attack. In *Proceedings of the First International Workshop on Peer-to-Peer Systems*, 2002.

[26] Duedil. Duedil – transparent, constructive feedback on your profile. http://www.duedil.com/, July 2010.

[27] Taher ElGamal. A public-key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Transactions on Information Theory*, IT-31(4):469472, 1985.

[28] Nurit Gal-Oz, Ehud Gudes, and Danny Hendler. A robust and knot-aware trust-based reputation model. In *Proceedings of the Joint iTrust and PST Conferences on Privacy, Trust Management and Security (IFIPTM 2008)*, 2008.

[29] Diego Gambetta. *Trust: Making and Breaking Cooperative Relatioins*, chapter Can We Trust Trust?, pages 213 – 237. Department of Sociology, University of Oxford, 2000.

[30] S. Goldwasser, S. Micali, and C. Rackoff. The knowledge complexity of interactive proof systems. *SIAM Journal on Computing*, 18(1):186208, 1989.

[31] Ehud Gudes, Nurit Gal-Oz, and Alon Grubshtein. Methods for computing trust and reputation while preserving privacy. In *Proceedings of the 23rd Annual IFIP WG 11.3 Working Conference on Data and Applications Security*, 2009.

[32] Minaxi Gupta, Paul Judge, and Mostafa Ammar. A reputation system for peer-to-peer networks. In *Proceedings of the 13th international Workshop on Network and Operating Systems Support for Digital Audio and Video*, 2003.

[33] Jochen Haller. A bayesian reputation system for virtual organizations. In *Negotiation, Auctions, and Market Engineering*, 2008.

[34] M. Ham and G. Agha. Ara: A robust audit to prevent free-riding in p2p networks. In *Proceedings of the Fifth IEEE International Conference on Peer-to-Peer Computing*, 2005.

[35] Omar Hasan, Lionel Brunie, and Elisa Bertino. k-shares: A privacy preserving reputation protocol for decentralized environments. In *Proceedings of the 25th IFIP International Information Security Confer-*

ence (SEC 2010), pages 253–264, Brisbane, Australia, September 2023 2010.

[36] Kevin Hoffman, David Zage, and Cristina Nita-Rotaru. A survey of attack and defense techniques for reputation systems. Technical Report CSD TR 07-013, Department of Computer Science, Purdue University, IN, USA, 2007.

[37] Jiangyi Hu and Mike Burmester. Lars a locally aware reputation system for mobile ad hoc networks. In *Proceedings of the 44th Annual Southeast Regional Conference*, Melbourne, Florida, USA, 2006.

[38] Audun Josang and Roslan Ismail. The beta reputation system. In *Proceedings of the 15th Bled Electronic Commerce Conference*, Bled, Slovenia, 2002.

[39] Audun Josang, Roslan Ismail, and Colin Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618 – 644, March 2007.

[40] Sepandar D. Kamvar, Mario T. Schlosser, and Hector GarciaMolina. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th International Conference on World Wide Web (WWW 2003)*, Budapest, Hungary, May 2003.

[41] Florian Kerschbaum. A verifiable, centralized, coercion-free reputation system. In *Proceedings of the 8th ACM workshop on Privacy in the electronic society (WPES'09)*. ACM, New York, NY, USA, 2009.

[42] Michael Kinateder and Siani Pearson. A privacy-enhanced peer-to-peer reputation system. In *Proceedings of the 4th International Conference on Electronic Commerce and Web Technologies*, 2003.

[43] Michael Kinateder, Ralf Terdic, and Kurt Rothermel. Strong pseudonymous communication for peer-to-peer reputation systems. In *Proceedings of the 2005 ACM symposium on Applied computing*, 2005.

[44] Raph Levien. *Attack-Resistant Trust Metrics (Chapter 5). Computing with Social Trust.* Springer London, 2008.

[45] Felix Gomez Marmol and Gregorio Martinez Perez. Security threats scenarios in trust and reputation models for distributed systems. *Computers & security*, 28:545 – 556, 2009.

[46] S. Marti and H. Garcia-Molina. Limited reputation sharing in p2p systems. In *Proceedings of the 5th ACM Conference on Electronic Commerce*, 2004.

[47] Chris Mitchell, editor. *Trusted computing.* Institution of Electrical Engineers, 2005.

[48] Jordi Nin, Barbara Carminati, Elena Ferrari, and Vicenc Torra. Computing reputation for collaborative private networks. In *Proceedings of the 33rd Annual IEEE International Computer Software and Applications Conference*, 2009.

[49] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford University, January 29 1998.

[50] Pascal Paillier. Public-key cryptosystems based on composite degree residuosity classes. In *Proceedings of the 17th International Conference on Theory and Application of Cryptographic Techniques*, 1999.

[51] Elan Pavlov, Jeffrey S. Rosenschein, and Zvi Topol. Supporting privacy in decentralized additive reputation systems. In *Proceedings of the Second International Conference on Trust Management (iTrust 2004)*, Oxford, UK, 2004.

[52] Siani Pearson and Boris Balacheff, editors. *Trusted Computing Platforms: TCPA Technology in Context.* Prentice Hall, 2003.

[53] Torben P. Pedersen. Non-interactive and information-theoretic secure verifiable secret sharing. In *Proceedings of the 11th Annual International Cryptology Conference on Advances in Cryptology*, 1991.

[54] Franziska Pingel and Sandra Steinbrecher. Multilateral secure cross-community reputation systems for internet communities. In *Proceedings of the Fifth International Conference on Trust and Privacy in Digital Business (TrustBus 2008)*, 2008.

[55] Paul Resnick and Richard Zeckhauser. Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system. *The Economics of the Internet and E-Commerce. Michael R. Baye, editor. Volume 11 of Advances in Applied Microeconomics*, pages 127–157, 2002.

[56] Stefan Schiffner, Sebastian Clau, and Sandra Steinbrecher. Privacy and liveliness for reputation systems. In *Proceedings of the Sixth European Workshop on Public Key Infrastructures, Services and Applications (EuroPKI'09)*, pages 209 – 224, 2009.

[57] Mudhakar Srivatsa, Li Xiong, and Ling Liu. Trustguard: Countering vulnerabilities in reputation management for decentralized overlay networks. In *Proceedings of the 14th International Conference on World Wide Web*, Chiba, Japan, 2005.

[58] Sandra Steinbrecher. Design options for privacy-respecting reputation systems within centralised internet communities. In *Security and Privacy in Dynamic Environments*, 2006.

[59] Ashwin Swaminathan, Renan G. Cattelan, Cherian V. Mathew, Ydo Wexler, and Darko Kirovski. Relating reputation and money in on-line markets. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, 2009.

[60] Unvarnished. Unvarnished – community-contributed reviews for business professionals. http://www.getunvarnished.com/, July 2010.

[61] Marco Voss, Andreas Heinemann, and Max Muhlhauser. A privacy preserving reputation system for mobile information dissemination networks. In *Proceedings of the First International Conference on Security and Privacy for Emerging Areas in Communications Networks (SECURECOMM)*, 2005.

[62] Yao Wang and Julita Vassileva. Bayesian network trust model in peer-to-peer networks. In *Agents and Peer-to-Peer Computing*, 2004.

[63] Andrew Whitby, Audun Josang, and Jadwiga Indulska. Filtering out unfair ratings in bayesian reputation systems. In *Proceedings of the Workshop on Trust in Agent Societies, at the Autonomous Agents and Multi Agent Systems Conference (AAMAS2004)*, New York, July 2004.

[64] Haifeng Yu, Michael Kaminsky, Phillip B. Gibbons, and Abraham Flaxman. Sybilguard: Defending against sybil attacks via social networks. In *Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM'06)*, pages 267–278. ACM Press, September 2006.

[65] Lan Yu. A reputation system for bittorrent peer-to-peer file-sharing networks. Master's thesis, University of Wollongong, Australia, 2006.

[66] Runfang Zhou and Kai Hwang. Powertrust: A robust and scalable reputation system for trusted peer-to-peer computing. *IEEE Transactions on Parallel and Distributed Systems*, 18(4):460–473, April 2007.

[67] Cai-Nicolas Ziegler and Georg Lausen. Spreading activation models for trust propagation. In *Proceedings of the 2004 IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE '04)*, 2004.