Synergies of different reputation systems: challenges and opportunities

Le-Hung Vu, Thanasis G. Papaioannou and Karl Aberer
School of Computer and Communication Sciences
École Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
{*lehung.vu, thanasis.papaioannou, karl.aberer*}*@epfl.ch*

***Abstract*—Reputation is a well established means to determine trustworthiness in online systems in various contexts, e.g. online transactions, product recommendation, e-mail spam fighting, etc. However, typically these reputation systems are "closed" outside of the community: the set of participants, their possible actions, their evaluation and the mechanism to derive trust evaluations are predetermined in the system design. Therefore, existing information is hardly reused and emerging online communities have a "cold" start regarding trustworthiness. In this paper, we discuss the various opportunities that arise by combining reputation information from different communities and provide a detail discussion on related challenges, namely identification, mapping of reputation semantics, contextual distance, reputation disclosure (dis)incentives and privacy. For example, the critical issue of identification can more effectively be dealt with based on entity-matching and the social structure of different systems. Furthermore, we argue and theoretically prove that even naïve combinations of reputation values from different communities can result in a system capable of detecting misbehavior more effectively than the individual reputation mechanisms themselves under certain conditions on reputation semantics.**

***Keywords*-online communities, trust inference, open systems, identity equivalence mapping**

I. Introduction

Reputation is a well established means to determine trustworthiness in online systems. A large number of reputation-based trust systems have been proposed, as comprehensively reviewed in [1]–[3]. Also, a number of commercial examples exists in various contexts, e.g. online transactions (eBay[1], ePier[2], iKarma[3]), product recommendation (Amazon, ePinions[4]), or e-mail spam fighting (TrustedSource, Trend Micro Email [5]). These systems calculate reputation values based on previous actions of the participants and probably on referrals to estimate future behavior of participants or expected quality. However, *typically these reputation systems are "closed"*: very little information about the user communities and the reputation mechanisms being used are exchanged with other systems. For example, the set of participants in most cases is known only within one system. In addition, possible actions of users, the evaluation of their behaviors, and the mechanism to derive trust from such user interactions are predetermined in the system design. Hence, each system exploits the history of users' actions in a closed environment. This is in stark contrast to the open nature of the Internet where no firm system boundaries and stable specifications of actions and information exist.

Nowadays, new online communities dynamically emerge in various contexts through the exchange of views and services among systems. The trustworthiness of the members of these communities is evaluated from scratch (i.e. all members are considered equally untrustworthy, "cold start"). However, this is not *socially optimal*, as the members of these communities are often also members of existing communities, possibly employing different virtual identities, and their trustworthiness in these contexts has already been evaluated.

A reputation system in a closed community is able to recognize specific untrustworthy behavior patterns within that context, given the structure of the community, the structure of the reputation system, and the history of actions in the context. However, people actively participate in various online communities, such as social networks (e.g. LinkedIn[6], myExperiment[7], Facebook[8]), online transaction environments or product review sites. *Reputation and behavior of a user in a system may well be related to his reputation and behavior in another*. Combining reputation information for the same members from different contexts may be more effective a mechanism for identifying *additional patterns* of malicious behavior within each individual context. Thus, reputation systems should take advantage of not only diverse information sources in the Internet, such as social networks, recommender systems, content sharing systems, or semantic search engines, but also other reputation systems.

In this paper, we investigate the various issues related to reusing existing reputation information from different contexts, namely identification, mapping of reputation semantics, contextual distance, reputation disclosure (dis)incentives and privacy. For example, there is a cost

[1]www.ebay.com
[2]www.epier.com
[3]www.ikarma.com
[4]www.epinion.com
[5]www.trendsecure.com/portal/en-US/tools/security_tools/emailid
[6]www.linkedin.com
[7]www.myexperiment.org
[8]www.facebook.com

associated to integrating reputation information from many systems that has to be compared to the benefit from opening them up. We also identify several opportunities for doing so, such as the ones described above.

The remainder of this paper is organized as follows: in the next section, we motivate the opening and exchanging reputation information among different systems and present possible opportunities for doing so. We then summarize the most relevant initiatives and related work in Section III. In Section IV, we prove, based on a generic model, that the combination of reputation information is beneficial for the case of an emerging community from two existing reputation systems. Section V identifies and gives a detailed discussion of various research challenges related to reusing reputation information and combining existing reputation systems. Section VI concludes the paper.

II. Motivation and Opportunities

In this section, we investigate various benefits of opening-up reputation systems, exchanging reputation information, and combining different systems and various opportunities for doing so. The combination of various reputation systems is expected to lead to a combination of benefits even beyond of them individually achieved by each system for the following reasons:

1. Trustworthiness information from one system can be used as prior information to learn user behaviors in another. Therefore, "cold start" of trust estimation can be avoided in an emerging community.
2. The estimation of trustworthiness in emerging communities does not solely rely on pre-specified reputation data from existing ones and can take advantage of diverse information sources in the Internet, such as social networks, community platforms, recommender systems, content sharing systems or semantic search engines.
3. Shared but hidden assumptions of communities on values and evaluations of actions can be discovered by reputation aggregation across communities that otherwise would be hard to discover and specify.
4. In dynamically formed emerging communities, self-organizing community processes form an agreement on the instantiation of the reputation model for a given context. These processes involve identifying and structuring sources of reputation data, extracting and agreeing on behaviors of interest (in particular specify malicious behavior) and their evaluation, and establishing shared evaluation metrics. The obvious source for building reputation metrics is relevance feedback from participants to determine categories of behavior and their evaluation. In this process, a community norm is compared and adapted based on the norms of other communities such as recommender systems, emergent semantic systems and web-ranking systems, and it can identify sub-communities with shared interests. To this end, a reputation system cannot be applied in these communities a priori, without hindering their formation. Emerging reputation systems can be introduced *a posteriori* and address trust issues as soon as community norms on expectable behavior and community assumptions have converged.
5. For identification, emerging reputation systems do not exclusively rely on system-internal identity but on Web-based entities, which are digital footprints in the Web and therefore not easily manipulated. Also, employing graph relations from different communities and complex network metrics (e.g. clique), emerging reputation systems have a better potential to counter identity attacks, which are a major problem in closed reputation systems. Therefore, aggregating reputation information from different contexts is expected to more effective in discovering collusive groups and in dealing more effectively with "Sybil" attacks [4].
6. Beneficial attributes of each trust computational model applied to different communities can be combined together when reputation information from different communities is utilized either inside each existing community or in an emerging one. Combined reputation information is expected to be more robust and resilient against attacks from users with various opportunistic and malicious behaviors. This opportunity is investigated in Section IV.

III. Related work

One of the first initiatives (2008) to opening reputation systems is that of the OASIS standardization forum. Specifically, an Open Reputation Management Systems (ORMS) technical committee [5] has been founded to provide the ability to use common XML data formats for representing reputation data and standard definitions of reputation scores. The standards will provide the means for understanding the relevance of a rating score within a given transaction across communities, but not the algorithms for computing the scores.

[6] introduces a software prototype to store ratings of users participating in different online-discussion fora. This work assumes complete trustworthiness of the identity provider and the users have full-control on the disclosure of actions and reputation information across communities.

The building of open reputation systems is also well-related to existing works on federated identity management [7]. For example, [8] defines a PKI framework for secure referral dissemination across communities. [9] proposes a generalized approach to model contextual environment of an agent and compute its trust values in a new context based on distance with the related situations. Emerging stan-

dards such as OpenID[9] and OpenAuth[10] enable the creation portable identities and allowing community managers to delegating access to users across system boundaries.

The use of Semantic Web RDF models as FOAF (Friend Of A Friend)[11] to weave social networks and SIOC (Semantically-Interlinked Online Communities)[12] to weave social activities such as blog comments could provide a complete interlinked graph on top of existing applications. Also, different community semantics can be expressed in meta-models such as OWL[13]. However, the adoption of common ontologies or schemas is still relatively slow. Moreover, the process of community building is highly dynamic. Emergent communities may dynamically form within existing ones with different interactions and objectives. As explained in [10], semantic interoperability should be viewed as an emergent phenomenon constructed incrementally, and its state at any given point in time depends on the frequency, the quality and the efficiency with which negotiations, such as those defined in [11], [12], can be conducted to reach agreements on common interpretations within the context of a given task. Specifically, in [12], semantic interoperability is addressed by means of local schema agreements between data sources through queries and gossiping of schema mappings. The quality of semantic mappings is gradually improved by a feedback algorithm until to finally achieve global agreement selecting the right data sources for schema translation.

Commercial solutions for aggregating online reputation information related to a person are available and are becoming increasingly popular, such as Online Reputation Monitor[14], Reputation Manager[15], or Reputation Defender[16]. However, these are mainly entity matching web search engines [13], as opposed to identity equivalence mapping explained in Section IV, that neither automatically check the accuracy of information nor aggregate it into a single trustworthiness metric.

To the best of our knowledge, our work is the first attempt to systematically study and analyze the benefits and challenges of exchanging and combining different reputation systems to improve system performance.

IV. Benefits of Combining Reputation Models-Preliminary Analysis

In this section, we present a generic model for the combination of two different reputation systems, in order to investigate the benefit of exchanging reputation information. Herein, we consider the two reputation systems being used in a similar application contexts. Therefore, ratings and reputation values in one system are well-defined and relevant (i.e. meaningful) to the other. An example of two such systems are the online e-commerce sites eBay[17] and ePier[18], which employ similar reputation mechanisms.

We want to evaluate whether such a combination of two reputation systems leads to any improvement of the robustness of each individual system against user misbehavior. Specifically, we measure the resilience of the combination of two systems, each of which is designed to be resilient against a certain misbehavior model. Our goal is to quantify the improvement (in terms of overall resilience) of the combined system as compared to each individual system and to measure the synergy effect of such a combination, if any.

We assume that community managers and initiators provide basic support for identifying and structuring relevant reputation data for mapping it into a representation with shared semantics. Therefore, we enable effective exchanges of reputation data and combination of information from different systems. For simplicity, each community manager is assumed to be trustworthy. This issue is discussed in detail in Section V.

A reputation-based trust system can be formally described as in Def. 1. As defined therein, we only consider *single-dimensional* reputation values with *common* and *well-defined* semantics.

Definition 1: A reputation system $\mathcal{R}$ is a 6-tuple $\mathcal{R} = \langle \mathcal{U}, id, \mathcal{K}, \mathcal{H}, \mathcal{T}, \mathcal{A} \rangle$ where:

- $\mathcal{U}$ is the set of usernames in the system. This information is usually public.
- $\mathcal{K}$ is the set of credentials to verify user identities. In most cases $\mathcal{K}$ is private, i.e., only known by the owning user and the community manager. Example credentials are emails, public keys, or even physical addresses.
- $id : \mathcal{U} \rightarrow \mathcal{K}$ is the system identity verification method (which may be done during the registration phase).
- $\mathcal{H} = \{h^u : u \in \mathcal{U}\}$ is historical performance data of all users.
- $\mathcal{T}$ is the set of completed transactions among users.
- $\mathcal{A} : \mathcal{U} \times \mathcal{T} \times \mathcal{H} \rightarrow \{trustworthy, undecided\}$ is an algorithm operating on $\mathcal{H}$ and outputs a value determining if a user $u \in \mathcal{U}$ is trustworthy for a given transaction $t \in \mathcal{T}$.

Let $\mathcal{B} = \{b^u, u \in \mathcal{U}\}$ be the overall behavior model of all users, where b^u is the behavior of a user u on the probability space. We make no assumption on each individual behavior b^u, thus the overall model $\mathcal{B}$ summarizes any possible malicious and opportunistic behaviors of any participants, including any collusive actions of a group.

The misclassification error rate of a reputation system

[9] www.openid.net
[10] www.openauth.net
[11] www.foaf-project.org
[12] www.sioc-project.org
[13] http://www.w3c.org/TR/owl-ref/
[14] http://reputation.distilled.co.uk/
[15] www.reputationmanager.com
[16] www.reputationdefender.com
[17] www.ebay.com
[18] www.epier.com

$\mathcal{R} = \langle \mathcal{U}, id, \mathcal{K}, \mathcal{H}, \mathcal{T}, \mathcal{A} \rangle$ given a user behavior model $\mathcal{B}$ is defined as the expectation that the system misclassifies behaviors of any user for a transaction (Def. 2). A lower misclassification error rate implies a higher resilience under malicious behaviors of users.

Definition 2: Let $0 \leq e(u, t \mid \mathcal{A}, \mathcal{B}, h^u) \leq 1$ be the probability the algorithm $\mathcal{A}$ misclassifies a user u as trustworthy for a transaction $t \in \mathcal{T}$, given the rating history of the user h^u and the overall behaviors of all users $\mathcal{B}$. The misclassification error rate of the system given the user behavior model $\mathcal{B}$ is $s = \frac{\sum_{u \in \mathcal{U}, t \in \mathcal{T}} e(u,t|\mathcal{A},\mathcal{B},h^u)}{|\mathcal{U}||\mathcal{T}|}$. The system resilience under the behavior model $\mathcal{B}$ is $r = 1 - s$.

Denote $h^u \parallel \Delta^u$ the integration of two historical performance data sets h^u and Δ^u of a user u. In this work, we limit our analysis to those algorithms $\mathcal{A}$ satisfying the following Hypothesis 1. Such a hypothesis is practically reasonable with a wide range of algorithms, since a better learning result is usually expected given more data.

Hypothesis 1: $e(u, t \mid \mathcal{A}, \mathcal{B}, h^u)$ is a monotonically decreasing function of $\mid h^u \mid$. In other words, $d(\Delta^u, \mathcal{A}, \mathcal{B}) = e(u, t \mid \mathcal{A}, \mathcal{B}, h^u \parallel \Delta^u) - e(u, t \mid \mathcal{A}, \mathcal{B}, h^u) \leq 0$

Let us combine two systems $\mathcal{R}_1$ and $\mathcal{R}_2$ with the same identity credential set $\mathcal{K}$. Denote $\mathcal{R}_i = \langle \mathcal{U}_i, id_i, \mathcal{K}, \mathcal{H}_i, \mathcal{T}_i, \mathcal{A}_i \rangle$, where $\mathcal{H}_i = \{h_i^u : u \in \mathcal{U}_i\}$, $i = 1, 2$, our goal is to combine $\mathcal{R}_1$ and $\mathcal{R}_2$ in order to achieve an overall resilience that is better or, at least, not worse than, each individual system under any misbehavior model $\mathcal{B}_j, j = 1, 2$. The integration of more than two systems can be done similarly.

For example, let $\mathcal{A}_1$ be a dishonesty detection algorithm designed to discover collusive groups of malicious users who consistently vote in favor for each other (possibly for a limited group size). Let $\mathcal{A}_2$ be another algorithm that detects periodical changes in behaviors of users, where period is finite. We seek to investigate if, combining the two systems $\mathcal{R}_1$ and $\mathcal{R}_2$, the final system is able to detect both collusively malicious and periodically changes of behaviors. Moreover, we want the resilience of the integrated system to be higher than the resilience of each individual system.

We shall show that such synergies can be achieved by combing the knowledge of two systems. Particularly, we exploit the fact that many users participate in both systems and their behaviors can be derived by aggregating their historical performance statistics in both systems. The detection of common participants of the two systems is related to the problems of alias detection and entity resolution in security and database research communities. For example, in [14], the authors managed to identify a large number of user identities from anonymized Flickr and Twitter data sets by analyzing their relationships with others with a very high accuracy, even if the two data sets are not strongly overlapped.

Basically, such an *identity equivalence mapping* can be done based on many information sources: First, many users use the same credentials in both system that uniquely identify them from the others (a.k.a. their digital footprints). For example, a user may use the same e-mail address to register as participants in the two systems. Other typical credentials can be detected from similarity in IP source addresses, correlations in posting timestamps, etc. An example solution to automatically detect such identity equivalences from various information sources is presented in [15]. Second, even if users use different credentials, equivalence among their usernames in two systems can be derived by analyzing the (trust) relationships among these users with the others. For example, methods to look for matching hidden patterns and structural stenography among users in the trust graphs of two systems [14], [16] can be applicable. Figure 1 illustrates the possibilities of such an identity equivalence mapping. Suppose three users u, v, w forms a collusion group in the first system $\mathcal{R}_1$, and other users u', x, y form another collusion in $\mathcal{R}_2$. By combining the two systems, algorithm $\mathcal{A}_1$ can be used in both communities to detect the two collusive groups effectively. Since u and u' actually use the same credential, e.g., the same email for identity verification, they are revealed to be associated to the same user. We can then estimate that v, w is related to x, y with certain probability.

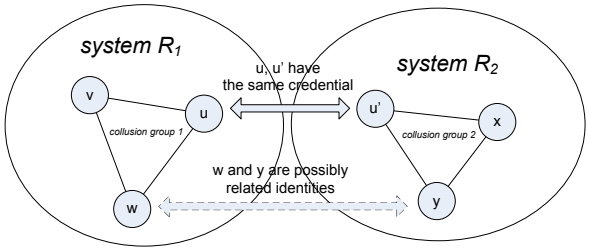

Figure 1. Inferring identity equivalence in the two reputation systems based on known knowledge and relationship analysis.

In this paper, we do not investigate specific solutions to this problem, but we assume the availability of an appropriate identity equivalence mapping mechanism for the integration of the two reputation systems. We define as $\tau = \{u \in \mathcal{U}_1 : \exists v \in \mathcal{U}_2, id_1(u) = id_2(v)\}$ the set of usernames with the same identity credential in both systems. Similarly we can define $\tau' = \{u \in \mathcal{U}_2 : \exists v \in \mathcal{U}_1, id_1(u) = id_2(v)\}$, and $\mid \tau \mid = \mid \tau' \mid$. Def. 3 formally introduces the notion of identity equivalence mapping between two user communities.

Definition 3: An identity equivalence mapping between the user communities of two systems $\mathcal{R}_1$ and $\mathcal{R}_2$ is defined as:

$$ide : \mathcal{U}_1 \times \mathcal{U}_2 \times \tau \rightarrow [0, 1] \quad (1)$$

The function $ide(u, v, \tau)$ gives the probability that user $u \in \mathcal{U}_1$ is the same as user $v \in \mathcal{U}_2$, knowing the relationship among users and the initial set of equivalent identities τ.

We denote as τ^* the set of users that can be reliably detected as common in both system by identity equivalence

mapping given τ. Formally, $\tau^* = \{u \in \mathcal{U}_1 : \exists v \in \mathcal{U}_2, ide(u,v,\tau) = 1\}$, and presumably $\tau^* \supseteq \tau$.

Similarly, we have $\tau'^* = \{u \in \mathcal{U}_2 : \exists v \in \mathcal{U}_1, ide(u,v,\tau') = 1\}$, where $\tau'^* \supseteq \tau'$, and $\mid \tau^* \mid = \mid \tau'^* \mid$ (see Figure 2).

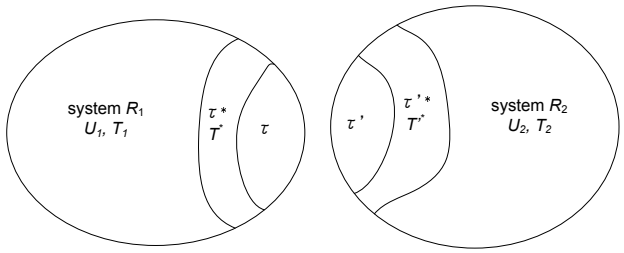

Figure 2. Combination of two reputation systems

Given the above notations, we introduce and analyze the resilience of the following naïve combination of the two reputation systems (Def. 4).

Definition 4: The naïve combination of two reputation systems $\mathcal{R}_i, i = 1,2$ is a reputation system $\mathcal{R} = \langle \mathcal{U}, id, \mathcal{K}, \mathcal{H}, \mathcal{T}, \mathcal{A} \rangle$ defined by:

- $\mathcal{U} = \mathcal{U}_1 \cup \mathcal{U}_2 - \tau^*$.
- $\mathcal{H} = \mathcal{H}_1 \parallel \mathcal{H}_2$ is the combination of historical data from $\mathcal{R}_1$ and $\mathcal{R}_2$. Specifically: $\mathcal{H} = \{h_1^u \parallel h_2^u : u \in \tau^*\} \cup \{h_1^u : u \in \mathcal{U}_1 - \tau^*\} \cup \{h_2^u : u \in \mathcal{U}_2 - \tau'^*\}$.
- $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2$
- $\mathcal{A} = \mathcal{A}_1 \circ \mathcal{A}_2$ is the combination of two algorithms $\mathcal{A}_1$ and $\mathcal{A}_2$. Under $\mathcal{A}$, a user u is evaluated as trustworthy iff given the same history of u, each $\mathcal{A}_1$ and $\mathcal{A}_2$ independently evaluates u as trustworthy.

Let $s_{ij} = \frac{\sum_{u \in \mathcal{U}_i, t \in \mathcal{T}_i} e(u,t|\mathcal{A}_i,\mathcal{B}_j,h_i^u)}{|\mathcal{U}_i||\mathcal{T}_i|}$, be the misclassification error rate of $\mathcal{R}_i$ given that the user behavior is $\mathcal{B}_j$, where $i,j \in \{1,2\}$. Suppose that $\mathcal{A}_i$ is designed to be more robust against the behavior model $\mathcal{B}_i$ than to the model $B_j, j \neq i$, it follows that $s_{ii} \leq \min\{s_{ij}, i \neq j\}$, where $i,j \in \{1,2\}$.

Denote $\overline{s}_i = \frac{\sum_{u \in \mathcal{U}, t \in \mathcal{T}} e(u,t|\mathcal{A}_1 \circ \mathcal{A}_2, \mathcal{B}_i, h_1^u \| h_2^u)}{|\mathcal{U}||\mathcal{T}|}$ the misclassification error rate of the naïve combination of two systems (Def. 4) given the user behavior model $\mathcal{B}_i, i = 1,2$. Proposition 1 gives us the estimation of the resilience of the combined system.

Proposition 1: Let $\mid \mathcal{U}_2 \mid = \alpha \mid \mathcal{U}_1 \mid$, $\mid \mathcal{T}_2 \mid = \beta \mid \mathcal{T}_1 \mid$, and $\mid \tau^* \mid = \gamma \mid \mathcal{U}_1 \mid$, where $0 \leq \alpha, \beta, \gamma \leq 1$. An upper-bound of the worst-case misclassification error of the combined system is given by:

(a) $\overline{s}_1 \leq \frac{s_{11} + \alpha\beta s_{21}}{(1+\alpha-\gamma)(1+\beta)}$.

(b) $\overline{s}_2 \leq \frac{s_{22} + \alpha\beta s_{12}}{(1+\alpha-\gamma)(1+\beta)}$.

Proof: Due to symmetry, we only need to prove (a). We note that under any user behavior model $\mathcal{B}_j, j = 1,2$, Def. 2 and Def. 4 give us:

$$e(u,t \mid \mathcal{A}_1 \circ \mathcal{A}_2, \mathcal{B}_j, h^u) = e(u,t \mid \mathcal{A}_1, \mathcal{B}_j, h^u) e(u,t \mid \mathcal{A}_2, \mathcal{B}_j, h^u) \quad (2)$$

From Hypothesis 1:

$$d(\delta^u, \mathcal{A}_i, \mathcal{B}_j) = e(u,t \mid \mathcal{A}_i, \mathcal{B}_j, h_k^u \parallel \delta^u) - e(u,t \mid \mathcal{A}_i, \mathcal{B}_j, h_k^u) \leq 0 \quad (3)$$

where $i,j,k \in \{1,2\}$.

Let $\mathcal{T}^* \subseteq \mathcal{T}_1$ (resp. $\mathcal{T}'^* \subseteq \mathcal{T}_2$) be the set of transactions where estimates of behavior of users in τ^* (resp. τ'^*) are made, the total misclassification error TME by the combined systems is given by:

$$\begin{aligned} TME \triangleq & \sum_{u \in \mathcal{U}, t \in \mathcal{T}} e(u,t \mid \mathcal{A}_1 \circ \mathcal{A}_2, \mathcal{B}_1, h_1^u \parallel h_2^u) = \\ & \sum_{u \in \tau^*, t \in \mathcal{T}^*} e(u,t \mid \mathcal{A}_1, \mathcal{B}_1, h_1^u \parallel h_2^u) e(u,t \mid \mathcal{A}_2, \mathcal{B}_1, h_1^u \parallel h_2^u) + \\ & \sum_{u \in \mathcal{U}_1 - \tau^*, t \in \mathcal{T}_1 - \mathcal{T}^*} e(u,t \mid \mathcal{A}_1, \mathcal{B}_1, h_1^u) e(u,t \mid \mathcal{A}_2, \mathcal{B}_1, h_1^u) + \\ & \sum_{u \in \tau'^*, t \in \mathcal{T}'^*} e(u,t \mid \mathcal{A}_1, \mathcal{B}_1, h_1^u \parallel h_2^u) e(u,t \mid \mathcal{A}_2, \mathcal{B}_1, h_1^u \parallel h_2^u) + \\ & \sum_{u \in \mathcal{U}_2 - \tau'^*, t \in \mathcal{T}_2 - \mathcal{T}'^*} e(u,t \mid \mathcal{A}_1, \mathcal{B}_1, h_2^u) e(u,t \mid \mathcal{A}_2, \mathcal{B}_1, h_2^u) \end{aligned} \quad (4)$$

On the other hand, considering reputation systems separately, the sum SME of their misclassification error is given by:

$$\begin{aligned} SME \triangleq & \sum_{u \in \tau^*, t \in \mathcal{T}^*} e(u,t \mid \mathcal{A}_1, \mathcal{B}_1, h_1^u) \\ & + \sum_{u \in \mathcal{U}_1 - \tau^*, t \in \mathcal{T}_1 - \mathcal{T}^*} e(u,t \mid \mathcal{A}_1, \mathcal{B}_1, h_1^u) \\ & + \sum_{u \in \tau'^*, t \in \mathcal{T}'^*} e(u,t \mid \mathcal{A}_2, \mathcal{B}_1, h_2^u) \\ & + \sum_{u \in \mathcal{U}_2 - \tau'^*, t \in \mathcal{T}_2 - \mathcal{T}'^*} e(u,t \mid \mathcal{A}_2, \mathcal{B}_1, h_2^u) \end{aligned} \quad (5)$$

Since $0 \leq e(.) \leq 1$, each term in Eq. (4) is less than the corresponding term in Eq. (5). Using Eqs. (2),(3), we then have:

$$\begin{aligned} TME - SME \leq & \sum_{u \in \tau^*, t \in \mathcal{T}^*} d(h_2^u, \mathcal{A}_1, \mathcal{B}_1) \quad (6) \\ & + \sum_{u \in \mathcal{U}_1 - \tau^*, t \in \mathcal{T}_1 - \mathcal{T}^*} 0 \\ & + \sum_{u \in \tau'^*, t \in \mathcal{T}'^*} d(h_1^u, \mathcal{A}_2, \mathcal{B}_1) \\ & + \sum_{u \in \mathcal{U}_2 - \tau'^*, t \in \mathcal{T}_2 - \mathcal{T}'^*} 0 \\ = & \sum_{u \in \tau^*, t \in \mathcal{T}^*} d(h_2^u, \mathcal{A}_1, \mathcal{B}_1) \\ & + \sum_{u \in \tau'^*, t \in \mathcal{T}'^*} d(h_1^u, \mathcal{A}_2, \mathcal{B}_1) \\ \triangleq & \Phi(\tau^*, \mathcal{T}^*, \tau'^*, \mathcal{T}'^*) \leq 0 \quad (7) \end{aligned}$$

The left-hand size of (4) can be rewritten as:

$$TME = (\mid \mathcal{U}_1 \mid + \mid \mathcal{U}_2 \mid - \mid \tau^* \mid)(\mid \mathcal{T}_1 \mid + \mid \mathcal{T}_2 \mid)\overline{s}_1 \quad (8)$$

We also have:

$$\begin{aligned} SME &= \mid \mathcal{U}_1 \mid\mid \mathcal{T}_1 \mid s_{11} + \mid \mathcal{U}_2 \mid\mid \mathcal{T}_2 \mid s_{21} \\ &= \mid \mathcal{U}_1 \mid\mid \mathcal{T}_1 \mid (s_{11} + \alpha\beta s_{21}) \end{aligned}$$

Thus (a) follows naturally. ■

Suppose $s_{ii} > 0$ and let $s_{ji} = \delta_i s_{ii}, i,j \in \{1,2\}, j \neq i$, where $\delta_i \geq 1$. Given Proposition 1, it is apparent that the combination of the two systems results in a better system if and only if $\overline{s}_i \leq s_{ii}, i = 1,2$. This is equivalent to the following condition:

$$f_i = \frac{(1+\alpha-\gamma)(1+\beta)}{1+\alpha\beta\delta_i} > 1, i = 1,2 \quad (9)$$

The condition (9) is satisfied if and only if:

$$1 < \delta_i < \frac{\alpha+\beta+\alpha\beta-(1+\beta)\gamma}{\alpha\beta} \text{ and } \gamma < \frac{\alpha+\beta}{1+\beta} \quad (10)$$

for $0 < \alpha, \beta < 1$.

The combined system has a benefit factor of $f_i = \frac{(1+\alpha-\gamma)(1+\beta)}{1+\alpha\beta\delta_i}$ under the user behavior model B_i compared to the individual system. Figure 3 shows the benefit factor for some example α, β, and γ values, i.e., a lower bound on the expected effectiveness improvement by combining different reputation systems.

Thus, the benefit factor reaches the maximal value for $\delta_i \to 1, i = 1, 2$. This condition is equivalent to $s_{11} \approx s_{21}$ and $s_{12} \approx s_{22}$, or that the two systems have approximate resilience under any behavior model $\mathcal{B}_1$ and $\mathcal{B}_2$. In other words, *it is best to combine two systems with comparable performance against different behavior models to achieve the highest synergy among them*.

From Eq. (7), one can also derive a tighter bound between $\overline{s}_1$ and s_{11}, s_{21} than the one of Proposition 1. This new bound depends on the function $\Phi(\tau^*, \mathcal{T}^*, \tau'^*, \mathcal{T}'^*)$, which can be estimated numerically given certain assumptions on the shape of the error function $e(.)$ (Def. 1). In such a case, it it apparent that the resilience of the combined system (represented by $\overline{s}_1, \overline{s}_2$) is even better. The synergy achieved in this situation is dependent on the set of common users τ^* and related transactions $\mathcal{T}^*$. The thresholds of τ^* and τ'^* to ensure the combination is meaningful and effective, as clearly shown, are $\mid \tau^* \mid = \mid \tau'^* \mid > 0$.

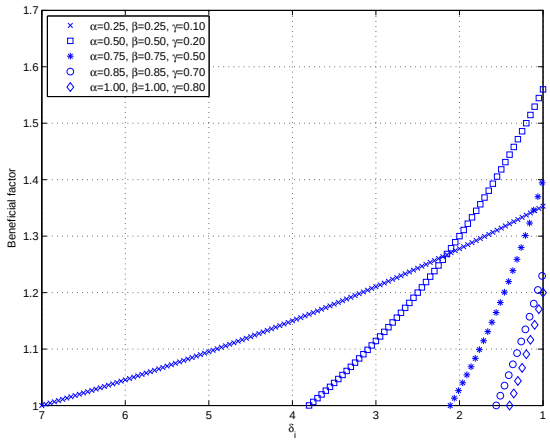

Figure 3. Benefit factor f_i vs. $\delta_i, i = 1, 2$

In summary, the total benefit (i.e, overall system resilience improvement under a mixture of behaviors) achieved by exchanging data and combining two reputation systems is contributed by two main factors: (1) The combined accuracy of two misbehavior detection algorithms, and (2) the effectiveness of the mechanisms for the reliable discovery of common users τ^* that enables effective reuse of user historical performance data for better evaluation of their behaviors.

V. Open Issues and Research Challenges

The possible benefits of opening up various reputation systems and combing them for synergic benefits are not without fundamental challenges that are not addressed in the previous analysis. In this section we give a detailed discussion of these open issues, and correspondingly identify many interesting related research questions deserving further work. Issues are presented in order of relevance and importance in our opinion.

Incentives for openness: Aggregating reputation within a community is a costly process. Furthermore, reputation information can be considered as a feature of the community and it is a public good for the members of the community. Therefore, it is not straightforward that community managers have the incentive to disclose reputation information to other communities, especially if the latter are competitive ones. This information is considered a business secret. For example, *eBay.com*[19] is unlikely to share reputation data with another e-trading system[20]. Community managers may have the incentive to improperly manipulate reputation information. To this end, a trustworthiness metric should be attached to each community regarding the accuracy of reputation information reported. Also, incentives for accurate reputation information exchange have to be provided, i.e. exchange on a reciprocative basis or buying/selling reputation information [17]. To a larger extent, automatic protocols for fair data exchange among different systems may be required.

The disclosure of information on identities and their relations are also of high importance. Moreover, raw transaction data could also be useful. Usually, online communities do not provide any inside information to outsiders for free. For example, Facebook's user profile data are valuable information to market researchers and thus can not be given for free by Facebook. Peering agreements among communities to share reputation-related information could create *externalities* for both communities by increasing their credibility to the users. This scenario particularly makes sense for federated small communities that are thus able to calculate trustworthiness more effectively over a larger user and transaction base and alleviate the "cold start" problem. Therefore, smaller communities are expected to be more willing to reciprocate. On the other hand, large well-established communities are not expected to be willing to disclose reputation-related data for free. Economic incentives can be provided to them by means of payments to provide reputation, identity or raw transaction data. However, in that case, competition among reputation-related information providers would arise. The accuracy of reputation or identity information, the size

[19] www.ebay.com
[20] e.g. www.ricardo.ch

of the community, the number of competing providers and the number of clients would determine actual prices. A federated group of many small communities can eventually be a strong reciprocative partner for a large system. For example, eBay may eventually lose members if ricardo.ch exchanges data with eBay's competitor alibaba.com[21].

Privacy issues: The Internet exposes personal information without providing any incentive for accountability. This information is subject to aggregation on-demand by various commercial systems, such as Spokeo.com. However, systems for aggregating such personal information may thus use unreliable information sources and even fake data, which may severely affect the reputation of a person. Improving reliability of information sources in emergent reputation systems may be a substantial aid in fighting malicious publication of personal information, disfamation attacks and contribute to better protection of personal privacy and integrity. However, the aggregation of even accurate reputation information and its linkage to a person may raise privacy concerns for the members of online communities. To this end, individual actions and interactions inside a community should not be linkable to real identities, but only to pseudonyms. Moreover, only relevant reputation information should be communicated across communities.

Identity equivalence mapping methods: One of the most critical short-comings of closed reputation systems is the possibility of whitewashing behaviors by acquiring cheap identities. In an emergent reputation system, the virtual identities of the participants can be related to existing contexts such as social networks or Semantic Web entities. However, how are identities resolved when integrating reputation systems from different communities? First, certain online communities may employ membership based on real-life attribute credentials. For example, ricardo.ch is an example of an auction site using physical addresses as a means of verification of user identities. Some social networks may even employ real identities or real life addressing for membership. Second, users may employ the same credentials across several communities. Third, indirect identity resolution techniques by analyzing correlated patterns of actions across different contexts to disambiguate the identity of participants can be well applicable. As entities in the Web constitute digital footprints, they cannot be easily created or changed and entity management in the Semantic Web (e.g. Friend-Of-A-Friend) allows increasingly relating information from widely spread resources to a subject of trust evaluation [13]. Another indirect way for identity resolution would be to combine information from different trust and rating graphs. This is analogous to the use of social network analysis. It has been proved in [14], [16] that if only the a limited number of links are known in an anonymous social network, it is possible to discover the true identities of all other nodes of the network. Also, employing graph relations from different communities and complex network metrics (e.g. clique), it is possible to discover collusive groups and deal more effectively with "Sybil" attacks.

Accuracy of identification and incentives provided: However, entity matching (i.e. identity resolution) among different contexts may be inaccurate. In this case, reputation and trust graph information is improperly mapped introducing some noise into trustworthiness estimation. Therefore, the incentives to users for acceptable behavior may get distorted. The relation between the accuracy of entity matching and the resulting incentives for users after aggregation of reputation information from different communities has to be investigated. This is partially answered in Section IV. Also, the aggregation of reputation information across communities and the subsequent trust evaluations also influence the dynamics inside communities. For example, a member evaluated as untrustworthy by other communities will be regarded as less trustworthy inside a community where it behaves cooperatively. As a result, the member's dominant strategy in the latter community is not to cooperate. It thus requires much more effort from the particular member than others to be regarded as trustworthy in an emerging community.

Reputation semantics, community structure, and system dynamics: Reputation within a certain context is calculated based on the specific metrics of reputation within that context (i.e. how is good reputation defined, how is rated), the specific computational trust model in use [18], the community structure and dynamics (i.e. interactions), the community norms on what is good or bad behavior (i.e. reputation semantics) and evaluation criteria. Past transactions may be rated or not, by one or both transacted parties, and with quantitative or qualitative feedback. The feedback messages are aggregated based on different trust computational models that may also weight the significance of the ratings by the credibility of the raters and take into account other factors, such as the transaction context factor (i.e. size and kind of transaction) and the community context factor (e.g. common incentives or beliefs) [19]. Rating, reputation, credibility metrics and the metrics of any other factor for reputation calculation and their semantics may have to be shared across different communities for reputation transferability.

To this end, OASIS standards [5] will provide the means for representing and understanding the relevance of a reputation score across communities, but not the algorithms for computing the scores. However, the context of a community also has to be effectively modeled and described. The semantic distance of the contexts has to be calculated in terms of relativeness or usefulness (as opposed to lexicographical, linguistic or physical) distance between their attributes [20]. Also, different communities have different assumptions and structure. The way that their members interact, their common

[21]www.alibaba.com

knowledge, their beliefs and expectations are different. As a result, even the same physical entity, being member of different communities may act differently. For example, consider a member of eBay that is reluctant of giving negative feedback fearing provider's retaliation, while in Amazon, the same person is very strict in her evaluation and mostly provides negative book reviews. Therefore, the assumptions and the structure of each community have to be taken into account when aggregating reputation from different contexts. For the description of the context, the context structure and its assumptions a shared model language that is expressive enough is needed.

Different communities may have different norms on which behavior is acceptable or not. Therefore, a physical entity with consistent behavior across communities may have high reputation in one community and low reputation in another. The evaluation criteria are also affected by the community norms but also from the context structure and community assumptions.

Metrics for evaluating reputation accuracy and effectiveness against malicious behavior: A trust computational model is able to identify certain patterns of misbehavior, but based on the specific characteristics of the model and the membership policy of the community, it may also have certain vulnerabilities, such as the incorporation of inaccurate or biased feedback. Several approaches for dealing with inaccuracy of feedback within one context are described in Section III. In many online communities members employ pseudonyms that cannot be mapped to real identities. As a result, the members of these communities may employ multiple virtual identities and strategically manipulate reputation information, e.g. by promoting themselves or their friends and by demoting their competitors. This form of malicious behavior is referred to as "Sybil" attack and some approaches that deal with it within a context are described in Section III. Accuracy of reputation information within a context has to efficiently estimated and its significance into reputation aggregation across contexts to be appropriately adjusted.

Integrating raw transaction data or reputation information: When integrating ratings and trust values rather than raw transaction data, the underlying semantics of generating these rating have to be taken into account. It is interesting that, depending on the trust evaluation model employed, the same reputation data can lead to quite diverse evaluations [21]. Furthermore, the dynamics and the structure of a certain community have to taken into account, e.g. kind of transactions, frequency, etc. Also, the credibility of feedback for raw transactions and the reliability assigned to a community for the reputation information provided is very important.

Integration cost vs. benefits: Aggregation of reputation information across different communities requires identity resolution, collection of reputation information and trustworthiness estimation taking into account the aforementioned issues. Therefore, as it involves considerable communication and processing latency, it is not considered as feasible to be done on demand. Also, it is very costly in terms of communication to continuously exchange reputation information with other communities for all of their members or to continuously run crawlers for this purpose. A hybrid approach combining proactive crawling over other communities for the reputation information of selected members of the community and on-demand reputation information requests for some members to other communities seems to be a more appropriate solution.

VI. Conclusion

This paper presents a first step towards the systematic identification of the benefits and the opportunities of opening up different reputation systems and exchanging information across them. We have analyzed and quantified (based on a simple model) possible benefits of a combined system. We have found that synergies of different reputation systems can be achieved from two main factors. The first source of performance improvement comes from the combination of reputation estimation algorithms with different accuracy under different misbehavior models. The second comes from potential discovery of common participants in the two systems. This discovery enables the effective reuse and integration of reputation information from many systems to evaluate user behaviors more accurately. We also provide an extensive discussion of many challenges related to the issue of opening-up reputation systems. Addressing these issues requires dedicated solutions for several interesting research questions across many research communities.

References

[1] Y. Wang and K.-J. Lin, "Reputation-oriented trustworthy computing in e-commerce environments," *IEEE Internet Computing*, vol. 12, no. 4, pp. 55–59, 2008.

[2] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decis. Support Syst.*, vol. 43, no. 2, pp. 618–644, 2007.

[3] J. Golbeck, "Trust on the world wide web: A survey," *Foundations and Trends in Web Science*, vol. 1, no. 2, pp. 131–197, 2006.

[4] J. R. Douceur, "The sybil attack," in *Peer-To-Peer Systems: First International Workshop, Iptps 2002, Cambridge, Ma, USA, March 7-8, 2002, Revised Papers*. Springer, 2002.

[5] Http://www.oasis-open.org/committees/orms.

[6] F. Pingel and S. Steinbrecher, "Multilateral secure cross-community reputation systems for internet communities," in *TrustBus '08: Proceedings of the 5th international conference on Trust, Privacy and Security in Digital Business*. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 69–78.

[7] R. Bhatti, E. Bertino, and A. Ghafoor, "An integrated approach to federated identity and privilege management in open systems," *Commun. ACM*, vol. 50, no. 2, pp. 81–87, 2007.

[8] A. Gutscher, "A trust model for an open, decentralized reputation system," in *Proceedings of the Joint iTrust and PST Conferences on Privacy Trust Management and Security (IFIPTM 2007)*, 2007.

[9] M. Rehák and M. Pechoucek, "Trust modeling with context representation and generalized identities," in *CIA*, 2007, pp. 298–312.

[10] K. Aberer, P. Cudré-Mauroux, A. M. Ouksel, T. Catarci, M. S. Hacid, A. Illarramendi, V. Kashyap, M. Mecella, E. Mena, and E. J. Neuhold, "Emergent semantics principles and issues," in *Database Systems for Advances Applications (DASFAA 2004), Proceedings*. Springer, March 2004, pp. 25–38.

[11] K. Aberer, P. Cudré-Mauroux, and M. Hauswirth, "The chatty web: Emergent semantics through gossiping," in *Proceedings of WWW*, May 2003.

[12] K. Aberer and M. Hauswirth, "Start making sense: The chatty web approach for global semantic agreements," *Journal of Web Semantics*, vol. 1, p. 2003, 2003.

[13] W. Shen, X. Li, and A. Doan, "Constraint-based entity matching," in *Proceedings of the AAAI*, July 2005.

[14] A. Narayanan and V. Shmatikov, "De-anonymizing social networks," *IEEE Security and Privacy (to appear)*, 2009. [Online]. Available: \url{http://randomwalker.info/social-networks/}

[15] P. Cudré-Mauroux, P. Haghani, M. Jost, K. Aberer, and H. de Meer, "idMesh: Graph-Based Disambiguation of Linked Data Engines," in *WWW'09: Proceedings of the 18th International World Wide Web conference*, Madrid, Spain, 2009.

[16] L. Backstrom, C. Dwork, and J. Kleinberg, "Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography," in *WWW '07: Proceedings of the 16th international conference on World Wide Web*. New York, NY, USA: ACM, 2007, pp. 181–190.

[17] S. Tadelis, "What's in a name? reputation as a tradeable asset," *American Economic Review*, vol. 89, no. 3, pp. 548–563, June 1999.

[18] L.-H. Vu and K. Aberer, "Effective usage of computational trust models in rational environments," Tech. Rep. LSIR-REPORT-2008-007, 2008. [Online]. Available: http://infoscience.epfl.ch/search?recid=125277&of=hd

[19] L. Xiong and L. Liu, "PeerTrust: Supporting reputation-based trust for peer-to-peer electronic communities." *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 7, pp. 843–857, 2004.

[20] J. F. Roddick, K. Hornsby, and D. de Vries, "A unifying semantic distance model for determining the similarity of attribute values," in *ACSC '03: Proceedings of the 26th Australasian computer science conference*. Darlinghurst, Australia, Australia: Australian Computer Society, Inc., 2003, pp. 111–118.

[21] Z. Despotovic and K. Aberer, "P2P reputation management: Probabilistic estimation vs. social networks," *Journal of Computer Networks, Special Issue on Management in Peer-to-Peer Systems: Trust, Reputation and Security*, vol. 50, no. 4, pp. 485–500, March 2006. [Online]. Available: http://dx.doi.org/10.1016/j.comnet.2005.07.003