Establishing Network Reputation via Mechanism Design*

Parinaz Naghizadeh Ardabili and Mingyan Liu

Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, Michigan, 48109-2122
{naghizad,mingyan}@umich.edu

Abstract. In any system of networks, such as the Internet, a network must take some measure of security into account when deciding whether to allow incoming traffic, and how to configure various filters when making routing decisions. Existing methods tend to rely on the quality of specific hosts in making such decisions, resulting in mostly reactive security policies. In this study we investigate the notion of reputation of a *network*, and focus on constructing mechanisms that incentivizes the participation of networks to provide information about themselves as well as others. Such information is collected by a centralized *reputation agent*, who then computes a *reputation index* for each network. We use a simple mechanism to demonstrate that not only a network has the incentive to provide information about itself (even though it is in general not true), but also that this information can help decrease the estimation error.

Keywords: Mechanism Design, Network Reputation, Incentives.

1 Introduction

This paper studies the following mechanism design problem: in a distributed multi-agent system where each agent possesses beliefs (or perceptions) of each other, while the truth about an agent is only known to that agent itself and it may have an interest in withholding the truth, how to construct mechanisms with the proper incentives for agents to participate in a collective effort to arrive at the correct perceptions of all participants without violating privacy and self-interest.

Our main motivation lies in the desire to enhance network security through establishing the right quantitative assessment of the overall security posture of different networks at a global level; such a quantitative measure can then be used to construct sophisticated security policies that are proactive in nature, which are distinctly different from current solutions that typically tackle specific security problems. Such quantitative measure can also provide guidance to networks' human operators to more appropriately allocate resources in prioritizing tasks – after all, the health of a network is very much a function of the due diligence of its human administrators.

* The work is partially supported by the NSF under grant CIF-0910765 and CNS-121768, and the U.S. Department of Commerce, National Institute of Standards and Technology (NIST) Technology Innovation Program (TIP) under Cooperative Agreement Number 70NANB9H9008.

V. Krishnamurthy et al. (Eds.): GameNets 2012, LNICST 105, pp. 47–62, 2012.

Consider a system of inter-connected networks. Each network has access to statistics gleaned from inbound and outbound traffic to a set of other networks. From these statistics it can form certain opinions about the quality or "cleanliness" of these other networks, and actions are routinely taken based on such opinions. For instance, network administrators may choose to block a high percentage of inbound traffic from a network observed to send out a large number of spams. Such peer network-network observations are often incomplete – a network does not get to see the entire traffic profile of another network – and can be biased. Thus two networks' view of a common third network may or may not be consistent.

The true quality of a network ultimately can only be known to that network itself, though sometimes a network may not have or choose to use the resources needed to obtain this knowledge. It is however not necessarily in the network's self-interest to truthfully disclose this information: a network has incentive to inflate other's perception about itself. This is because this perceived high quality often leads to higher visibility and less blocked outbound traffic from this network. Similarly, a network may or may not wish to disclose truthfully what it observes about others for a variety of privacy considerations. On the other hand, it is typically in the interest of all networks to have the *correct* perception about other networks. This is because this correct view of others can help the system administrator determine the correct security configurations.

In this paper we set out to examine the validity and usefulness of a *reputation system*, where a central *reputation agent* solicits input from networks regarding their perceptions of themselves and others, and computes a *reputation index* for each network as a measure/indicator of the health or security posture of a network. These reputation indices are then broadcast to all networks; a network can in turn combine such reputation information with its local observations to take proactive measures to maintain "good" reputation and/or improve its own reputation over time, and take proactive measures to protect themselves against networks with "bad" reputations. The ultimate goal of this type of architecture is to improve global network security, which has been championed by and is gaining support from network operators' organizations, see e.g., [10].

The design and analysis of such a system must observe two key features. The first is that participation in such a system is completely voluntary, and therefore it is critical for the system to adopt mechanisms that can *incentivize* networks to participate. The second is that networks may not report truthfully to the reputation agent even if they choose to participate in such a collaborative effort, and therefore it is crucial for any mechanism adopted by the system to either provide the right incentive to induce truth revelation, or be able to function despite untruthful input from networks.

It should be noted that a wide variety of systems have been developed to determine *host reputation* by monitoring different types of data. Darknet monitors [2], DNS sensors [1], scanning detection, firewall logs [3], web access logs, and ssh brute force attack reports are all examples of systems that can report on hosts that have engaged in potentially suspicious behavior. The most commonly used host reputation systems are related to determining illegitimate email messages or SPAM.

A wide range of different organizations such as SPAMHAUS [12], SpamCop [6], Shadowserver [14], and Barracuda [11], independently operate their own reputation lists, which are largely generated by observing unauthorized email activity directed at monitored spamtraps. In addition, organizations such as Team Cymru [8], Shadowserver, and Damballa [7] generate similar reputation lists by analyzing malware or even DNS activity. There is however a significant difference between assessing individual hosts' reputation vs. defining reputation as a notion for a network. Host reputation lists by themselves cannot directly be used in developing a consistent security policy due to the dynamic nature of host addresses.

Besides the security context, there has been a large volume of literature on the use of reputation in peer-to-peer (P2P) systems and other related social network settings. Specifically, a large population and the anonymity of individuals in such social settings make it difficult to sustain cooperative behavior among self-interested individuals [5]. Reputation has thus been used in such systems as an incentive mechanism for individuals to *cooperate* and behave according to a certain *social norm* in general [15], and to reciprocate in P2P systems in particular [4,13,9]. While the focus of social network studies is on the effect of changing reputation has on individuals, the focus of our study in its present form is on how to make network reputation an accurate representation of a network's security posture. Accordingly, our emphasis is on how to incentivize participation from networks, while user participation in a P2P system is a given (i.e., by default reputation only applies to an active user already in a P2P system).

Our main findings are summarized as follows. We propose a reputation mechanism which induces a network to participate in the collective assessment of its own reputation. We first show that for two networks (Section 3), a network's participation can result in a higher mean estimated reputation and at the same time lower estimation error, thus benefiting both itself and the system. This remains true even if the observations of the other network is biased. We further show in Section 4 that these results extend to the case of multiple interacting networks.

2 The Model, Main Assumptions, and Preliminaries

2.1 The Model

Consider a system of K inter-connected networks, denoted by $N_1, N_2, \cdots, N_K$. Network N_i's overall health condition is described by a quantity r_{ii}, which will also be referred to as the *true* or *real reputation* of N_i, or simply the *truth*. We will assume without loss of generality that these true quantities are normalized, i.e., $r_{ii} \in [0, 1]$, for all $i = 1, 2, \cdots, K$.

There is a central *reputation agent*, who solicits and collects a vector $(X_{ij})_{j \in K}$ of *reports* from each network N_i. It consists of *cross-reports* $X_{ij}, i, j = 1, 2, \cdots, K, j \neq i$, which represent N_i's assessment of N_j's quality, and *self-reports* $X_{ii}, i = 1, 2, \cdots, K$, which are the networks' *self-advertised* quality measure disclosed to the reputation agent. The reputation agent's goal is to compute a *reputation index* denoted by $\hat{r}_i$, which is an estimate of r_{ii} for each network

N_i using a certain mechanism with the above inputs collected from the networks. This index/estimate will then be used by peer networks to regulate their interactions with N_i.

2.2 Assumptions

We assume that each network N_i is aware of its own conditions and therefore knows r_{ii} precisely, but this is in general its *private information*. While it is technically feasible for any network to obtain r_{ii} by closely monitoring its own hosts and traffic, it is by no means always the case due to reasons such as resource constraints.

We also assume that a network N_i can sufficiently monitor inbound traffic from network N_j so as to form an estimate of N_j's condition, denoted by R_{ij}, based on its observations. However, N_i's observation is in general an *incomplete* view of N_j, and may contain error depending on the monitoring and estimation technique used. We will thus assume that R_{ij} is described by a Normal distribution $\mathcal{N}(\mu_{ij}, \sigma_{ij}^2)$, which itself may be unbiased ($\mu_{ij} = r_{jj}$) or biased ($\mu_{ij} \neq r_{jj}$). We will further assume that this distribution is known to network N_j (a relaxation of this assumption is also considered later). The reason for this assumption is that N_j can closely monitor its outbound traffic to N_i, and therefore may sufficiently infer how it is perceived by N_i. On the other hand, N_i itself may or may not be aware of the distribution $\mathcal{N}(\mu_{ij}, \sigma_{ij}^2)$.

A *reputation mechanism* specifies a method used by the reputation agent to compute the reputation indices, i.e., how the input reports are used to generate output estimates. We assume the mechanism is common knowledge among all K participating networks.

A participating network N_i's objective is assumed to be characterized by the following two elements: (1) it wishes to obtain from the system as accurate as possible a reputation estimate $\hat{r}_j$ on networks N_j *other than itself*, and (2) it wishes to obtain as high as possible an estimated reputation $\hat{r}_i$ on *itself*. It must therefore report to the reputation agent a carefully chosen $(X_{ij})_{j \in K}$, using its private information r_{ii}, its knowledge of the distributions $(R_{ji})_{j \in K \setminus i}$, and its knowledge of the mechanism, to increase (or inflate) as much as possible $\hat{r}_i$ while keeping $\hat{r}_j$ close to r_{jj}. The reason for adopting the above assumption is because, as pointed out earlier, accurate assessment of other networks' security posture can help a network configure its policies appropriately, and thus correct perception of other networks is critical. On the other hand, a network has an interest in inflating its own reputation so as to achieve better visibility and less traffic blocked by other networks, etc. Note that these two elements do not fully define a network's preference model (or utility function). We are simply assuming that a network's preference is increasing in the accuracy of others' reputation estimate and increasing in its own reputation estimate, and that this is public knowledge[1].

[1] How the preference increases with these estimates and how these two elements are weighed remain the network's private information and do not factor into the present analysis.

Note also that the objective assumed above may not capture the nature of a *malicious* network, who may or may not care about the estimated perceptions about itself and others. Recall that our basic intent through this work is to provide reputation estimate as a quantitative measure so that networks may adopt and develop better security policies and be incentivized to improve their security posture through a variety of tools they already have. Malicious networks are not expected to react in this manner. On the other hand, it must be admitted that their participation in this reputation system, which cannot be ruled out as malicious intent may not be a priori knowledge, can very well lead to skewed estimates, thereby rendering the system less than useful. The hope is that a critical mass of non-malicious networks will outweigh this effect, but this needs to be more precisely established and is an important subject of future study.

2.3 Candidate Mechanisms and Rationale

One simple mechanism that can be used by the reputation agent is to take the estimate $\hat{r}_i$ to be the average of the cross-reports X_{ji} and the self-report X_{ii}. It can be easily seen that in this case, N_i will always choose to report $X_{ii} = 1$, and thus the self-reports will bear no information. The mechanism can be modified to take the average of only the cross-reports $(X_{ji})_{j \in K \setminus i}$ as the estimate. If cross-reports are unbiased, then $\hat{r}_i$ can be made arbitrarily close to r_{ii} as the number of networks increases. We will later take the mean absolute error of this mechanism, which we will refer to as the *averaging mechanism*, as a benchmark in evaluating the performance of other mechanisms.

An alternative to completely ignoring N_i's self-report is to induce or incentivize N_i to provide *useful* information in its self-report even if it is not the precise truth r_{ii}. With this in mind, a good mechanism might on one hand convince N_i that it can help contribute to a desired, high estimate $\hat{r}_i$ by supplying input X_{ii}, while on the other hand try to use the cross-reports, which are estimates of the truth r_{ii}, to assess N_i's self-report and threaten with punishment if it is judged to be overly misleading.

Also, note that it is reasonable to design a mechanism in which N_i's cross-reports are not used in calculating its own reputation estimate. By doing so, we ensure that the cross-reports are reported truthfully[2]. To see why this is the case, note that by its cross-reports N_i can now only hope to increase its utility by altering $\hat{r}_j$. Now N_i's best estimate of r_{jj} is R_{ij}, which it knows will be used as a basis for the estimate $\hat{r}_j$. On the other hand, due to its lack of knowledge of r_{jj}, N_i can't use a specific utility function to see how it can strategically choose X_{ij} so as to increase its utility. By this argument, for the rest of the paper we will assume that the cross-reports are reported truthfully, and that this is common knowledge.

It is worthwhile to emphasize that the above reasoning on truthful cross-reports derives from accounting for the *direct* effect of the cross-reports on the

[2] This is conceptually similar to not using a user's own bid in calculating the price charged to him in the context of auction, a technique commonly used to induce truthful implementation.

final estimates. One might argue that a network could potentially improve its *relative* position by providing false cross-reports of other networks so as to lower their reputation indices, i.e., it can make itself look better by comparison. A close inspection of the situation reveals, however, that there is no clear incentive for a network to exploit such *indirect* effect of their cross-reports either.

One reason is that the proposed reputation system is not a *ranking* system, where making other entities look worse would indeed improve the standing of oneself. The reputation index is a value normalized between $[0, 1]$, a more or less absolute scale. It is more advisable that a network tighten its security measures against *all* networks with low indices rather than favor the highest-indexed among them.

But more importantly and perhaps more subtly, badmouthing another network is not necessarily in the best interest of a network. Suppose that after sending a low cross-report X_{ij}, N_i subsequently receives a low $\hat{r}_j$ from the reputation agent. Due to its lack of knowledge of other networks' cross-reports, N_i cannot reasonably tell whether this low estimate $\hat{r}_j$ is a consequence of its own low cross-report, or if it is because N_j was observed to be poor(er) by other networks and thus $\hat{r}_j$ is in fact reflecting N_j's true reputation (unless a set of networks *collude* and jointly target a particular network). This ambiguity is against N_i's interest in obtaining accurate estimates of other networks; therefore bashing is not a profitable deviation from truthful reporting.

3 A Two-Network Scenario

3.1 The Proposed Mechanism

We start by considering only two networks and extend the result to multiple networks in the next section. We will examine the following way of computing the reputation index $\hat{r}_1$ for N_1, where ϵ is a fixed and known constant. The expression for $\hat{r}_2$ is similar, thus for the remainder of this section we will only focus on N_1.

$$\hat{r}_1(X_{11}, X_{21}) = \begin{cases} \frac{X_{21}+X_{11}}{2} & \text{if } X_{11} \in [X_{21} - \epsilon, X_{21} + \epsilon] \\ X_{21} - |X_{11} - X_{21}| & \text{if } X_{11} \notin [X_{21} - \epsilon, X_{21} + \epsilon] \end{cases} \tag{1}$$

In essence, the reputation agent takes the average of self-report X_{11} and cross-report X_{21} if the two are sufficiently close, or else punishes N_1 for reporting significantly differently. Note that this is only one of many possibilities that reflect the idea of weighing between averaging and punishing; for instance, we can also choose to punish only when the self-report is higher than the cross-report, and so on.

3.2 Choice of Self-report

As stated earlier, we assume N_1 believes N_2's cross-report is a sample of a random variable with distribution $X_{21} \sim \mathcal{N}(\mu, \sigma^2)$. As a result, the choice of

the self-report X_{11} is determined by the solution of the optimization problem $\max_{X_{11}} E[\hat{r}_1]$. Using (1), $E[\hat{r}_1]$ eventually simplifies to (with $F()$ and $f()$ denoting the cdf and pdf, respectively):

$$E[\hat{r}_1] = X_{11} + \tfrac{\epsilon}{2}(F(X_{11}+\epsilon) - 3F(X_{11}-\epsilon)) - \tfrac{1}{2}\int_{X_{11}-\epsilon}^{X_{11}+\epsilon} F(x)\mathrm{d}x - 2\int_{-\infty}^{X_{11}-\epsilon} F(x)\mathrm{d}x \ . \quad (2)$$

Taking the derivative with respect to X_{11} we obtain:

$$\frac{\mathrm{d}E}{\mathrm{d}X_{11}} = 1 + \frac{\epsilon}{2}[f(X_{11}+\epsilon) - 3f(X_{11}-\epsilon)] - \frac{1}{2}[F(X_{11}+\epsilon) + 3F(X_{11}-\epsilon)]. \quad (3)$$

We next re-write $\epsilon = a\sigma$; this expression of ϵ reflects how the reputation agent can limit the variation in the self-report using its knowledge of this variation σ [3]. Replacing $X_{21} \sim \mathcal{N}(\mu, \sigma^2)$ and $\epsilon = a\sigma$ in (3), and making the change of variable $y := \frac{X_{11}-\mu}{a\sigma}$ results in:

$$\frac{a}{\sqrt{2\pi}}(e^{-(\frac{a(y+1)}{\sqrt{2}})^2} - 3e^{-(\frac{a(y-1)}{\sqrt{2}})^2}) - \frac{1}{2}(\mathrm{erf}(\frac{a(y+1)}{\sqrt{2}}) + 3\mathrm{erf}(\frac{a(y-1)}{\sqrt{2}})) = 0 \ . \quad (4)$$

Therefore, if y solves (4) for a given a, the optimal value for X_{11} would be $X_{11}^* = \mu + a\sigma y$. Equation (4) can be solved numerically for a, resulting in Figure 1. It's interesting to see in that in Figure 1 we always have $y < 1$, and as a consequence $X_{11}^* < \mu + \epsilon$. This means that N_1 is choosing a self-report within its prediction of the acceptable range. Also note that this self-report is always positively biased, reflecting N_1's interest in increasing $\hat{r}_1$.

3.3 Value of Cross-Report and Self-report

We next examine how close the resulting reputation estimate $\hat{r}_1$ is to the real quality r_{11} by calculating the mean absolute error (MAE) and comparing it to that of the averaging mechanism; from this we further illustrate the roles and values of cross-report and self-report. We do this separately for two cases, where the cross-report comes from an unbiased distribution and a biased distribution, respectively. Note that in both cases the averaging mechanism for the two-network scenario reduces to taking the cross-report as the estimate, i.e. the averaging mechanism has an estimate of $E[X_{21}]$ for N_1.

Unbiased Cross-Report. We now compare the performance of (1) to the averaging mechanism.

[3] Note that we are assuming σ is known by the reputation agent as well as the networks. σ can be thought of as a measure of the variation of N_2's estimate, which depends on the nature of its observation and the algorithm it uses for the estimate. While this is not entirely an unreasonable assumption, it ultimately needs to be verified through analysis of real data.

Define $e_m := E[|\hat{r}_1 - r_{11}|]$ as the MAE of the mechanism described in (1) with $\epsilon = a\sigma$. As already derived, N_1's self-report is set to $X_{11}^* = \mu + a\sigma y$, where y solves (4) for a given a; N_2's cross-report X_{21} is set to R_{21} (truthful reporting); and R_{21} is assumed to be unbiased. With these assumptions, we find the following expression for e_m [4]:

$$\begin{aligned} e_m = &\frac{1}{2}\int_{\mu-ay\sigma}^{\mu+a(y+1)\sigma} xf(x)\mathrm{d}x - \frac{1}{2}\int_{\mu+a(y-1)\sigma}^{\mu-ay\sigma} xf(x)\mathrm{d}x \\ &- 2\int_{-\infty}^{\mu+a(y-1)\sigma} xf(x)\mathrm{d}x + ay\sigma + (\mu - ay\sigma)\ F(\mu - ay\sigma) \\ &+ (\mu + ay\sigma)\ \left(\frac{3}{2}F(\mu + a(y-1)\sigma) - \frac{1}{2}F(\mu + a(y+1)\sigma)\right). \end{aligned} \tag{5}$$

As seen in (5), e_m is a function of a. Thus we can optimize the choice of a by solving the problem $\min_a e_m$. Taking the derivative of (5) we get:

$$\begin{aligned} \frac{\mathrm{d}e_m}{\mathrm{d}a} = \frac{\sigma}{2}\Bigg(&\frac{a}{\sqrt{2\pi}}(e^{-\frac{(a(y+1))^2}{2}} - 3e^{-\frac{(a(y-1))^2}{2}}) + (ay + y')\Bigg(2 + \mathrm{erf}(\frac{ay}{\sqrt{2}}) + \\ &\frac{a}{\sqrt{2\pi}}(e^{-\frac{(a(y+1))^2}{2}} + 3e^{-\frac{(a(y-1))^2}{2}}) - \frac{1}{2}\Big(\mathrm{erf}(\tfrac{a(y+1)}{\sqrt{2}}) - 3\mathrm{erf}(\tfrac{a(y-1)}{\sqrt{2}})\Big)\Bigg)\Bigg). \end{aligned} \tag{6}$$

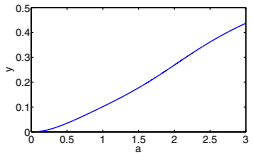

Fig. 1. Solution of (4): y vs. a

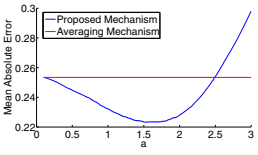

Fig. 2. Errors vs. a, $r_{11} = 0.75, \sigma^2 = 0.1$

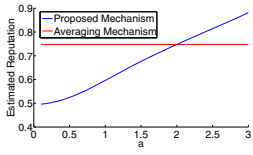

Fig. 3. Est. reputation vs. a, $r_{11} = 0.75, \sigma^2 = 0.1$

As seen in (6), the optimal choice of a does not depend on the specific values of μ and σ. Therefore, the same mechanism can be used for any set of networks. Equation (6) can be solved numerically, and is zero at two values: at $a = 0$, which indicates a local maximum, and at $a \approx 1.7$, where it has a minimum. This can been seen from Figure 2, which shows the MAE of the proposed mechanism compared to that of the averaging mechanism. Under the averaging mechanism the MAE is $E[|R_{21} - r_{11}|] = \sqrt{\frac{2}{\pi}}\sigma$. We see that for a large range of a values the mechanism given in (1) results in smaller estimation error. This suggests that N_1's self-report can significantly benefit the system as well as all networks other than N_1.

[4] The calculations here are possible if $y \leq \frac{1}{2}$, which based on Figure 1 is a valid assumption for moderate values of a.

We next examine whether there is incentive for N_1 to provide this self-report, i.e., does it benefit N_1 itself? Figure 3 compares N_1's estimated reputation $\hat{r}_i$ under the proposed mechanism to that under the averaging mechanism, in which case it is simply N_2's cross-report X_{21}, and $E[X_{21}] = \mu$ when unbiased.

Taking Figs 2 and 3 together, we see that there is a region, $a \in [2, 2.5]$ in which the presence of the self-report helps N_1 obtain a higher estimated reputation, while helping the system reduce its estimation error on N_1. This is a region that is mutually beneficial to both N_1 and the system, and N_1 clearly has an incentive to participate and provide the self-report.

Biased Cross-Report. We now turn to the case where the cross-report X_{21} comes from the biased distribution $\mathcal{N}(r_{11} + b, \sigma^2)$, where b is the bias term, a fact unknown to both N_2 and the reputation mechanism. We will thus assume that the mechanism used remains that given by (1) with the optimal value of a obtained previously.

First consider the case that N_1 is also not aware of the bias, and again chooses $X_{11}^* = r_{11} + ay\sigma$. The calculation of the error is the same, leading to (5). However, here F and f are those of the Normal distribution $\mathcal{N}(r_{11} + b, \sigma^2)$. Therefore, the new minimum error and the value of a where it occurs are different. Figure 4 shows the MAE for three different values of the bias. As seen from the figure, the error increases for $b = -0.1\sigma$, and decreases for $b = 0.1\sigma$ compared to the unbiased case. This is because for the negative bias, N_1 is not adapting its self-advertised reputation accordingly. This makes the mechanism operate mainly in the punishment phase, which introduces larger errors. For the small positive bias, however, the mechanism works mainly in the averaging phase, and the error is less than both the biased and unbiased cases. The latter follows from the fact that punishment phases happen more often in the unbiased case. Note however that for larger values of positive bias, the error will eventually exceed that of the unbiased case.

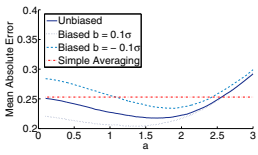

Fig. 4. MAE, biased cross-reports, bias not known

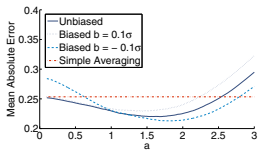

Fig. 5. MAE, biased cross-reports, bias known

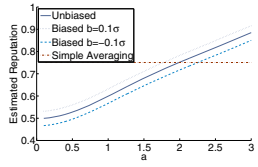

Fig. 6. Est. reputation, biased cross-reports, bias known

Next we consider the case where $X_{21} \sim \mathcal{N}(r_{11} + b, \sigma^2)$ as before but this bias is known to N_1. N_1 will accordingly adapt its self-report to be $X_{11}^* = r_{11} + b + ay\sigma$. Figure 5 shows a comparison in this case. The results show that the selected positive bias increases the error, while the negative bias can decrease the error compared to the unbiased case.

The assumption of a known bias has the following two intuitively appealing interpretations. The first is where N_1 has deliberately sent its traffic through N_2 in such a way so as to bias the cross-report. As expected, it's in the interest of N_1 to introduce a positive bias in N_2's evaluation of itself. If this is what N_1 chooses to do then arguably the mechanism has already achieved its goal of improving networks' security posture – after all, N_2 now sees a healthier and cleaner version of N_1 which is welcomed! The second case is where given the mechanism, N_2 knows that N_1 will introduce a positive bias in its self-report, and consequently counter-acts by sending a negatively biased version of its observation. To find the optimal choice for this deliberately introduced bias we proceed as follows. Define $\mu := r_{11} + b$. To see how the mean absolute error behaves, we find an expression for e_m at any given a.[5]

$$\begin{aligned} e_m = (\mu - r_{11} + ay\sigma) + \frac{1}{2}\int_{2r_{11}-\mu-ay\sigma}^{\mu+a(y+1)\sigma} xf(x)\mathrm{d}x - \frac{1}{2}\int_{\mu+a(y-1)\sigma}^{2r_{11}-\mu-ay\sigma} xf(x)\mathrm{d}x \\ - 2\int_{-\infty}^{\mu+a(y-1)\sigma} xf(x)\mathrm{d}x + (2r_{11} - \mu - ay\sigma)\ F(2r_{11} - \mu - ay\sigma) \\ + (\mu + ay\sigma)(\frac{3}{2}F(\mu + a(y-1)\sigma) - \frac{1}{2}F(\mu + a(y+1)\sigma))\ . \end{aligned} \tag{7}$$

where F and f are the cdf and pdf of the biased distribution. To find the value of b at which the error is minimized, we take the derivative of (7), resulting in:

$$\frac{\mathrm{d}e_m}{\mathrm{d}\mu} = 1 - 2F(2r_{11} - \mu - ay\sigma) = 0\ . \tag{8}$$

Solving (8) will show that for a given a, the MAE is minimized at $b^* = -\frac{ay\sigma}{2}$. As a result, the final reports sent by the two networks will be $X_{11}^* = r_{11} + \frac{ay\sigma}{2}$ and $X_{21}^* = R_{21} - \frac{ay\sigma}{2}$, which in turn increases the chance of having the mechanism operate in the averaging phase, thus decreasing the error.

As in the unbiased case, we also compare the estimated reputation $\hat{r}_1$ in this case to highlight that there is incentive for N_1 to provide self-report, shown in Figure 6. A comparison between Figs. 5 and 6 reflects the tradeoff between achieving a lower estimation error and helping N_1 achieve a higher estimated reputation. In the case of positive bias, even though N_1 benefits from providing a self-report for smaller values of a compared to the unbiased case, the system can use a more limited range of a to decrease MAE compared to the averaging mechanim. Similarly, larger values of a are required for incentivizing N_1's participation when the cross-report is negatively biased, while the MAE improvement is achieved for a larger range of a.

4 Extension to a Multi-network Scenario

We now consider the case with more than two participating networks. The proposed mechanism can be extended as follows. The reputation agent now receives

[5] The following calculations are for moderate values of bias $b \in [-ay\sigma, -ay\sigma + \frac{a\sigma}{2}]$.

more cross-reports on the basis of which it will judge N_i. In the simplest case, the agent can take the average of all the cross-reports to get $X_{0i} := \frac{1}{K-1}\Sigma_{j\in K\setminus i}X_{ji}$, and derive $\hat{r}_i$ using:

$$\hat{r}_i(X_{ii}, X_{0i}) = \begin{cases} \frac{X_{0i}+X_{ii}}{2} & \text{if } X_{ii} \in [X_{0i}-\epsilon, X_{0i}+\epsilon] \\ X_{0i} - |X_{ii} - X_{0i}| & \text{if } X_{ii} \notin [X_{0i}-\epsilon, X_{0i}+\epsilon] \end{cases} . \quad (9)$$

Another alternative is using a weighted version of the cross-reports in this mechanism. We defer this discussion to later in the section. For the mechanism defined in (9), we again have two cases, one where the cross-reports are unbiased, and one where they are biased. In the second case, we further distinguish between the cases where the bias itself is of a non-skewed distribution and where the bias distribution is skewed.

4.1 Unbiased Cross-Reports

We will assume $X_{ji} \sim \mathcal{N}(\mu_{ji}, \sigma^2_{ji})$, and that these distributions are independent. Thus X_{0i} also has a Normal distribution given by $\mathcal{N}(\frac{\Sigma_{j\in K\setminus i}\mu_{ji}}{K-1}, \frac{\Sigma_{j\in K\setminus i}\sigma^2_{ji}}{(K-1)^2})$. The optimization problem for N_i is the same as before resulting in $X^*_{ii} = \mu' + ay\sigma'$, with μ' and σ'^2 being the mean and variance of X_{0i}. Note that in this case the reputation agent is using $\epsilon = a\sigma'$.

If all cross-reports are unbiased, i.e., $\mu_{ji} = r_{ii}$, and $\sigma_{ji} = \sigma$, we have $X_{0i} \sim \mathcal{N}(r_{ii}, \frac{\sigma^2}{K-1})$. To find the optimal choice of a we will need to solve (6) again, with the only difference that σ is replaced by σ'. Therefore, the optimal choice of a, which is independent of the mean or variance of the reports, will be the same as before. This result can be verified in Figures 7 and 8, which show the MAE of collections of 3 and 10 networks respectively. Furthermore, as expected the error decreases as the number of networks increases in this case.

4.2 Biased Cross-Reports

Now assume that the cross-reports are biased and that the bias term itself comes from a Normal distribution. We re-write $X_{ji} = R_{ji} + B_{ji}$, where $R_{ji} \sim \mathcal{N}(r_{ii}, \sigma^2_{ji})$, and $B_{ji} \sim \mathcal{N}(b_{ji}, \sigma^2_{b,ji})$. Therefore, assuming independence, in general we have:

$$X_{0i} \sim \mathcal{N}(r_{ii} + \frac{\sum_{j\in K\setminus i} b_{ji}}{K-1}, \frac{\sum_{j\in K\setminus i}(\sigma^2_{ji} + \sigma^2_{b,ji})}{(K-1)^2}) . \quad (10)$$

Non-skewed Bias Distribution. If the bias distribution has zero mean ($b_{ji} = 0$) and all variance terms are the same: $\sigma_{ji} = \sigma$ and $\sigma_{b,ji} = \sigma_b$, then (10) is simplified to $X_{0i} \sim N(r_{ii}, \sigma''^2)$, where $\sigma''^2 = \frac{\sigma^2+\sigma^2_b}{K-1}$. The calculation of the optimal self-report is given by the same optimization problem as before, resulting in $X^*_{ii} = r_{ii} + ay\sigma''$. Figures 9 and 10 show the simulation results for $K = 3$ and $K = 10$ respectively. As expected, biased cross-reports result in larger error

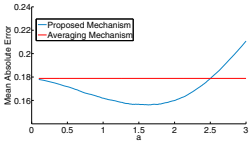

Fig. 7. MAE, 3 Networks, Unbiased Cross-Reports

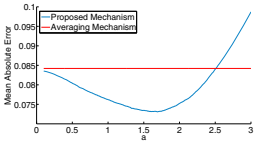

Fig. 8. MAE, 10 Networks, Unbiased Cross-Reports

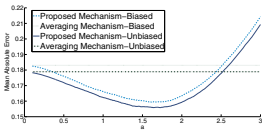

Fig. 9. MAE, 3 Networks, non-skewed bias distribution

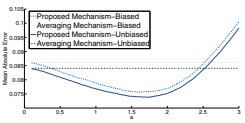

Fig. 10. MAE, 10 Networks, non-skewed bias distribution

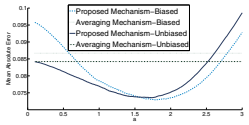

Fig. 11. MAE, 10 Networks, skewed bias distribution

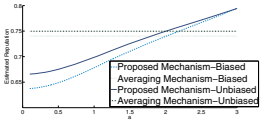

Fig. 12. Est. Reputation, 10 Networks, skewed bias distribution

compared to unbiased cross-reports: the fact that $\sigma'' > \sigma'$ in the unbiased case allows N_1 to introduce a larger inflation in its self-report, thus increasing the MAE in general.

Skewed Bias Distribution. If we assume that all bias terms are from the same distribution but this distribution is skewed itself, i.e. $B_{0i} \sim \mathcal{N}(b_{0i}, \sigma_b)$, then negatively biased cross-reports can result in lower MAE compared to a non-skewed bias distribution, while positively biased cross-reports can increase the error. Figure 11 verifies this property of the mechanism in a collection of 10 networks, and for a negative value of b_{0i}.

In all of the above cases, we need the range of a to be such that using the proposed mechanism is mutually beneficial for the system and the individual networks. Our numerical results show that, when cross-reports are unbiased, the values of a for which it is individually rational for a network to participate does not change as the number of networks increases. Also, this range remains unchanged if the cross-reports have a non-skewed bias distribution. In the case of skewed bias distribution a similar behavior as the two-network scenario is observed, where individual networks have more incentive to participate in the estimation of their own reputation when there is a positive bias in the cross-reports, and are less inclined to do so in the presence of a negative bias.

Figure 12 illustrates these results. As seen in the figure, for unbiaed cross-reports, the range for which networks are incentivized to participate is again roughly $a \in [2, 2.5]$ despite the increase in the number of networks. The figure also shows the effect of a choice of $b = -0.1\sigma$ for cross-reports with skewed bias. A careful study of this figure along with Figure 11 indicates that the same

tradeoff described in section 3 holds between minimizing error and providing incentive for participation.

4.3 Weighted Mean of Cross-Reports

So far, we have assumed the reputation agent takes a simple average of the cross-reports to judge the truthfulness of the self-report. Assume that as suggested earlier, the agent forms the weighted mean:

$$X_{0i} := \frac{\sum_{j \in K \setminus i} w_j X_{ji}}{\sum_{j \in K \setminus i} w_j} \tag{11}$$

where $\underline{w} := (w_j)_{j \in K \setminus i}$ is a vector of weights, also specified by the reputation agent. One reasonable choice for $\underline{w}$ could be a vector of previously computed reputations $\hat{r}_j$, with the goal of allowing the more reputable networks to have a higher influence on the estimate. We proceed by analyzing the performance of this alternative mechanism.

Unbiased Cross-Reports. Assume $X_{ji} \sim N(r_{ii}, \sigma_{ji}^2)$. By adopting this assumption, we focus on a scenario where all networks have an unbiased view of N_i, but potentially different accuracy as reflected by different values of σ_{ji}, with smaller variances corresponding to more precise estimates. Consequently, the weighted mean in (11) has a distribution $X_{0i} \sim N(r_{ii}, \frac{\sum_{j \in K \setminus i} {w_j}^2 \sigma_{ji}^2}{(\sum_{j \in K \setminus i} w_j)^2})$. Thus except for the change in the equivalent variance, the overall problem remains the same as the one discussed earlier[6]. Since an increased variance increases the MAE, in order to have a better estimate using the weighted average compared to the simple average, we would need $\frac{\sum_{j \in K \setminus i} {w_j}^2 \sigma_{ji}^2}{(\sum_{j \in K \setminus i} w_j)^2} \leq \frac{\sum_{j \in K \setminus i} \sigma_{ji}^2}{(K-1)^2}$.

In the special case $\sigma_{ji} = \sigma$, $\forall j$, the Cauchy-Schwarz inequality implies $\frac{\sum_{j \in K \setminus i} {w_j}^2}{(\sum_{j \in K \setminus i} w_j)^2} \geq \frac{1}{K-1}$, with equality at $w_j = w_0$, $\forall j$. This is true independent of the choice of $\underline{w}$, and therefore the weighted average will always have higher estimation error. Figure 13 shows this result for a random choice of the vector $\underline{w}$.

Next consider the case where σ_{ji}'s are different. Without lose of generality assume that the coefficients are normalized such that they sum to 1. In order to achieve lower estimation error, we want to choose $\underline{w}$ such that $\sum_{j \in K \setminus i} {w_j}^2 \sigma_{ji}^2 \leq \sum_{j \in K \setminus i} \frac{1}{(K-1)^2} \sigma_{ji}^2$. This rearrangement shows clearly that for the inequality to hold, it suffices to put more weight on the smaller σ_{ji}, i.e., more weight on those with more accurate observations. It follows that if more reputable networks (higher $\hat{r}_j$) also have more accurate observations (smaller σ_{ji}), then selecting weights according to existing reputation reduces the estimation error. Figure 14 shows the results for 3 networks when $\sigma_{31} < \sigma_{21}$, and the weights are chosen accordingly to be $w = (0.45, 0.55)$.

[6] In fact, using a simple average of cross-reports is a special case of this problem by using equal w_j and σ_{ji}.

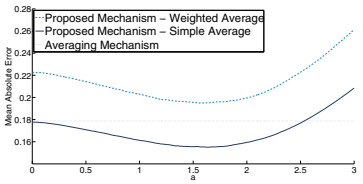

Fig. 13. MAE, 3 Networks, Weighted Averages, Equal Variances

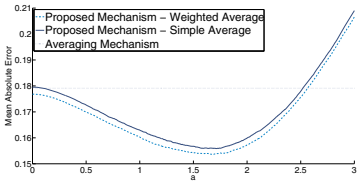

Fig. 14. MAE, 3 Networks, Weighted Averages, Different Variances

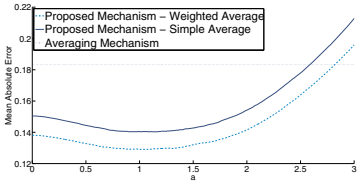

Fig. 15. MAE, 3 Networks, Weighted Averages, Skewed Bias

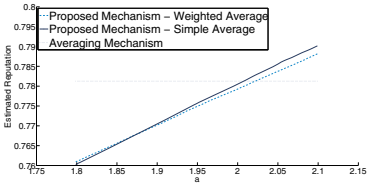

Fig. 16. Est. Reputation, 3 Networks, Weighted Averages, Skewed Bias

Biased Cross-Reports. Assume now $X_{ji} \sim N(r_{ii} + b_{ji}, \sigma_{ji}^2 + \sigma_{b,ji}^2)$. Then (11) results in $X_{0i} \sim N(r_{ii} + \frac{\sum_{j \in K \setminus i} w_j b_{ji}}{\sum_{j \in K \setminus i} w_j}, \frac{\sum_{j \in K \setminus i} {w_j}^2 (\sigma_{ji}^2 + \sigma_{b,ji}^2)}{(\sum_{j \in K \setminus i} w_j)^2})$. The case of equally distributed bias terms is very similar to before, and it will only add a bias term to the mean of the equivalent X_{01}. Therefore, we only focus on the case where b_{ji}'s are different.

In this case we have two ways of improving the result over the simple averaging. Following our previous discussion, putting more weight on the cross-reports that have smaller variances will decrease the final variance and thus the estimation error. On the other hand, if we put more weight on smaller bias terms, the overall bias will decrease. As already discussed in the beginning of this section, positively biased cross-reports increase the estimation error. Thus, having a smaller bias term will improve the MAE. Figure 15 shows the results for 3 networks, where N_3 has a better estimate than N_2, by which we mean both $0 < b_{31} < b_{21}$ and $\sigma_{31} < \sigma_{21}$. The weights are chosen such that $w_3 > w_2$.

Finally, we check networks' incentives under the weighted version of the mechanism. Based on our previous observation, we expect a similar tradeoff here as well: the lower MAE comes at the cost of the reduction in the range of a that makes the mechanism individually rational. This effect is illustrated in Figure 16.

5 Discussion and Conclusion

We demonstrated the feasibility of designing network reputation mechanisms that can incentivize networks to participate in the collective effort of determining

their health conditions by providing information about themselves and others. We showed that our mechanism can allow both the participants and the system to benefit. Furthermore, the mechanism remains robust even if we relax the assumption of unbiased initial estimation. As a byproduct of this analysis, we observed how once the mechanism is fixed, networks can improve the assessment even further by strategically choosing their cross-reports. We also verified that the same results hold as the number of participating networks increases.

This is only the first step toward building a comprehensive global reputation system; there remain many interesting and challenging problems to pursue.

To begin, the mechanisms proposed here (simple and weighted averages) are just two of many possible choices. In particular, it would be desirable to relax the assumption of having known variances, σ_{ij}^2, throughout the system, and see if it is possible to design alternative mechanisms that can achieve the same or better performance. Secondly, in practice it is possible for the reputation agent to obtain direct observations of its own as additional input to the estimation. This may allow us to relax the assumption that the cross-reports are truthful (though as we have argued this is a reasonable assumption in and by itself). Thirdly, it would be very interesting to analyze the effect of the presence of a small percentage of malicious networks as discussed in the paper.

At an architectural level, it would be of great interest to design a *distributed* mechanism without the need for a central reputation agent. One possibility is to follow a gossip-like procedure, where neighboring networks update their respective estimates using values provided by other networks through a similar averaging-punishment process to ensure that peer networks provide *useful* if not entire true information. It would be interesting to see what type of computation will lead to system-wide convergence to accurate estimates of the networks' health conditions.

References

1. Antonakakis, M., Perdisci, R., Dagon, D., Lee, W., Feamster, N.: Building a Dynamic Reputation System for DNS. In: 19th USENIX Security Symposium (August 2010)
2. Bailey, M., Cooke, E., Myrick, A., Sinha, S.: Practical Darknet Measurement. In: 40th Annual Conference on Information Sciences and Systems (March 2006)
3. DShield. How To Submit Your Firewall Logs To DShield (September 2011), `http://isc.sans.edu/howto.html`
4. Feldman, M., Lai, K., Stoica, I., Chuang, J.: Robust incentive techniques for peer-to-peer networks. In: ACM Conference on Electronic Commerce, pp. 102–111 (2004)
5. Hanaki, N., Peterhansl, A., Dodds, P., Watts, D.: Cooperation in evolving social networks. Management Science 53(7), 1036–1050 (2007)
6. Cisco Systems Inc. SpamCop Blocking List - SCBL (May 2011), `http://www.spamcop.net/`
7. Damballa Inc. Damballa Threat Reputation System (May 2011), `http://www.damballa.com/`

8. Team Cymru Inc. Malicious Activity Insight (May 2011), http://www.team-cymru.com/Services/Insight/
9. Kamvar, S., Schlosser, M.T., Molina, H.G.: The Eigentrust Algorithm for Reputation Management in P2P Networks. In: International Conference on World Wide Web, pp. 640–651 (2003)
10. Karir, M., Creyts, K., Mentley, N.: Towards Network Reputation - Analyzing the Makeup of RBLs. In: NANOGG52, Denver, CO (June 2011), http://www.merit.edu/networkresearch/papers/pdf/2011/NANOG52_reputation-nanog.pdf
11. Barracuda Networks. Barracuda Reputation Blocklist (May 2011), http://www.barracudacentral.org/
12. The SPAMHAUS project. SBL, XBL, PBL, ZEN Lists (May 2011), http://www.spamhaus.org/
13. Ravoaja, A., Anceaume, E.: STORM: A Secure Overlay for P2P Reputation Management. In: International Conference on Self-Adaptive and Self-Organizing Systems, pp. 247–256 (2007)
14. ShadowServer. The ShadowServer Botnet C&C List (May 2011), http://www.shadowserver.org/
15. Zhang, Y., van der Schaar, M.: Peer-to-Peer Multimedia Sharing based on Social Norms. Elsevier Journal on Signal Processing: Image Communication Special Issue on Advances in Video Streaming for P2P Networks (to appear)