Towards a Sybil-resilient reputation metric for P2P applications

Tien Tuan Anh Dinh, Mark Ryan
School of Computer Science, University of Birmingham, United Kingdom
{ttd,mdr}@cs.bham.ac.uk

Abstract

In this paper, we propose a reputation-based trust model for P2P applications and study it a security framework. This framework makes it easy to reason about the resilience of the reputation metrics against Sybil attacks. We show that using the PageRank reputation metric in a P2P system can make it very vulnerable from Sybil attacks. We then propose a new reputation metric based on the original PageRank, and show that it is more resilient against such attacks.

1. Introduction

In Peer-to-Peer (P2P) applications (content distribution, communication, online market, etc), *trust* has become more than just a desirable property, due to the existence of uncooperative and even malicious participants. Most problems caused by such peers can be eliminated if peers can establish trust among each other and use that to determine who is trustworthy enough to interact with. The computational model of trust has recently been studied in details. In this paper, we will use the model proposed by Mui [5], which has been widely used. In this model, trust can be inferred through *reputation*. Trust and reputation are two social metrics that are closely related and often used interchangeably. However, they are not computationally equivalent. Given $D_b = \{x_1, x_2, ..x_n\}$ as outcomes of past transactions that involved entity b (x_i has value of either 0 or 1 when the outcome is bad or good, respectively), the reputation of b, rep_b, is defined as a function of D_b, namely $rep_b = \theta(D_b)$. Trust of an entity a towards b is defined as the expectation that the next transaction between a and b will have a positive outcome, given the past behavior of b: $\tau_{ab} = Expc\{x_{ab}(n+1) = 1|D_b\}$. It can be seen that trust is always subjective, while reputation can be objective. Furthermore, reputation may be a good indicator of trust, especially when there is no other evidence. In P2P context, the lack of a central authority allows for *Sybil attacks* [2] on the reputation metric rep. In such attacks, the adversary obtains multiple identities (or *Sybils*) and use them to engineer his reputation. In a P2P-based online market application, for example, a Sybil attack can be carried out by an attacker creating multiple accounts, faking transactions and good feedback among those accounts.

In this paper, we propose a Sybil-resilient reputation model for P2P applications. The reputation metrics are studied in a model that contains a trusted, centralized component. We show that the reputation metric based on PageRank is very vulnerable to Sybil attacks. The degree of manipulation that is possible in P2P networks is much worse than that in Web networks. Another contribution is that we derive clustering information from the network and utilize that to develop a new reputation metric that is more resilient against Sybil attacks. This new metric also produces intuitive reputation values. Our reputation model for P2P is discussed in section 2. PageRank's vulnerability to Sybil attacks is shown in section 3. Our new reputation metric is discussed in section 4, followed by experimental studies that demonstrate its resilience against Sybil attacks. Further discussion and the conclusion are presented in section 6 and 7, respectively.

2. Reputation model for P2P

In our model, the building blocks for evaluating reputation are user feedback in the form of *transaction ratings*. A transaction could be an upload/download operation in content distribution applications, or a buying/selling interaction in an online P2P-based market. Entities involved in transactions give each other ratings based on their satisfaction. More formally, the system is modeled as a tuple (V, T, i, r, Rt). $V = \{1, 2, ...n\}$ is the set of peers in the system. $T = \{t_1, t_2, ..., t_s\}$ is the set of transactions. $i : T \mapsto V$ and $r : T \mapsto V$ are functions that return the *initiator* and *responder* of a transaction. $Rt : V \times T \mapsto R^+$ returns the rating that one peer gives to another, for a given transaction. $Rt(v, t)$ is defined iff $i(t) = v$ or $r(t) = v$. We impose that Rt returns positive values, meaning that negative feedback is not included in our model.

From this model, we derive a reputation graph $G = (V, E)$, in which V is the set of peers as above. E is the set

of edges: $(a, b) \in E$ if $\exists t \in T$ such that $i(t) = a \wedge r(t) = b$ or $i(t) = b \wedge r(t) = a$. Every edge has an associated weight, which is defined via a function $W : E \mapsto R^+$. Denote $T_{ab} = \{t \in T | i(t) = a \wedge r(t) = b\}$, which represents all transactions that a has with b. Then, $W((a, b)) = \frac{\sum_{(t \in T_{ab})} Rt(a, T_{ab})}{|T_{ab}|}$.

We utilize the ideal/real model [6], which is one of the fundamental technique used in security to define and analyze cryptographic protocols, to reason about the resilience of reputation metrics against Sybil attacks. More specifically, a reputation metric rep is studied in the ideal model, which consists of a *trusted, incorruptible party* and all peers interact directly to that party. The trusted party (TP) *collects* feedback from peers, then *computes* the metric rep. In the real model, there is no TP and peers interact directly with each other. A protocol P in the real model that evaluates the metric rep is at most as resilient as rep. The *realization* of such protocol P is at least as challenging as finding a good metric rep. In this paper, we focus on the latter. An obvious advantage of having such powerful TP is that we could, at the moment, ignore the computational problem and focus on the security (Sybil resilience) property of the reputation metric.

There exists many different reputation metrics. The original PageRank [7] is one of the *objective* reputation metrics, as it produces rankings that are agreed by all peers. Advogato [4] is one of the *subjective* reputation metrics, as the rankings are relative to a set of root nodes. Objective reputation metrics are more suitable for P2P systems than the subjective ones, because many peers do not know each other in advance, meaning that there may be no paths connecting them in the reputation graph. In this case, asymmetric metrics are clearly unsuitable, as they rely on the existence of such paths.

3. PageRank reputation metric

PageRank, based on Markov process model (or eigen-based), is the most well-known symmetric reputation metric. The PageRank model consists of a transition matrix T and a probability value ϵ, called the *jumping factor*. The reputation vector $\overrightarrow{rep}^T$ is defined as the stationary vector of a Markov process, which is the solution of the equation $\overrightarrow{rep}^T = (1 - \epsilon).\overrightarrow{rep}^T.T + \overrightarrow{\epsilon}^T$

Friedman *et al* [1] study how PageRank can be manipulated by Sybil attacks and show analytically that with respect to rank increase, one can achieve a dramatic improvement by adding more Sybils. PageRank is designed for directed (Web) graphs, but the reputation graph of a P2P system is, on the other hand, undirected. We experimented with 6064-node graphs to see whether such the difference in nature of the graphs affect PageRank's resilience against Sybil attacks. The Web graph is derived from actual links taken from the academic Web link database [1]. For the same number of nodes, we generated P2P graphs, whose degree distribution is power-law or has small-world property. We observe the attacker's ranks before and after the attack. Essentially, the improvement in reputation is considerably higher in P2P graphs than in Web graph. In a small-world graph, it seems that attacker can obtain the highest rank after the Sybil attack, regardless of his original rank. It was observed that the variance of the PageRank distribution of P2P graphs is smaller than that of Web graphs. As a consequence, a small increase in reputation value could result in noticeable increase in rank.

4. Cluster-based PageRank

The intuition behind our reputation metric comes from the observation that in a Sybil strategy, the attacker creates links with its Sybils in order to traps the random walk (used in the PageRank model) as long as possible within its region. Such a region controlled by the attacker would have many internal links and only a small number of external links (to the rest of the network). In other words, these regions forms *graph clusters*. We can see that clustering information of a graph can be very useful in detecting and dealing with Sybil attacks. In our new reputation metric, when the random walk arrives at a highly cohesive (clustered region), it adjusts the ϵ value to quickly jump out of that region and therefore avoid being trapped. We expect that rankings produced by the new metric to be reasonably close to those produced by the original PageRank metric.

Given a reputation graph G, our reputation metric, called cluster-based PageRank (or CPR) consists of a clustering algorithm CA, and two functions r and f. CA partitions G into k clusters $C = \{C_1, C_2, ..., C_k\}$. r evaluates the *density* metric d_i for a cluster C_i, i.e. $r : C \mapsto R^+$. f translates a density value d_i to a ϵ_i value, i.e. $f : R^+ \mapsto [0, 1]$. Two nodes in the same cluster are given the same value of ϵ_i.

We modify PageRank so that each node has a personalized value ϵ_i instead of a global value ϵ. Using the power method, reputation of a node i, rep_i can be evaluated through many iterations as follows:

$$rep_i^{k+1} = \sum_j (1 - \epsilon_j) \cdot M_{ji} \cdot rep_j^k + \frac{1}{n} \cdot \sum_j \epsilon_j \cdot rep_j^k \quad (1)$$

Any implementation of the algorithm needs to specify CA, r and f in order to derive the ϵ_i values. Graph clustering is a NP-complete, problem, but there are many approaches towards finding *good* clusters in graphs: multi-level, hyper-graph, circuit, spectral, ... In this paper, we use

[1] http://cybermetrics.wlv.ac.uk/database/

the Markov-based clustering algorithm (MCL) [9], which has been showed to produce good clusters for large graphs. The intuition behind MCL is that a random walk starting from a node i in a dense region will tend to stay in that region for a long time before moving to another region. Given a transition matrix M $(n \times n)$, the algorithm consists of two main operations: *expansion* and *inflation*. The expansion operation, $E_e(M)$, basically extends the current random walk to e extra hops, i.e. $E_e(M) = M^e$. The inflation operation with a parameter l, Γ_l, promotes the probability of jumping to a node in the local region: $\Gamma_l(M_{ij}) = \frac{(M_{ij})^l}{\sum_{k=0}^{n}(M_{kj})^l}$. These operations are alternated on M until the limit matrix T becomes *doubly idempotent*, or $E_e(T) = T$ and $\Gamma_l(T) = T$. The matrix T is extremely sparse and its structure is interpreted as clusters.

5. Experimental study

We study the resilience of our reputation metric against Sybil attacks with different reputation graphs. We use results from recent studies of P2P systems [8, 10] to generate graph topologies that are similar to those of real P2P systems. In particular, small-world graphs are generated using the Watts-Strogatz model and have the *clustering coefficient* value of 0.018. Power-law graphs are generated with the Eppstein model. Size (of both small-world and power-law graphs) is varied from 1,000 to 50,000 nodes.

In the attacker model, we assume only one Sybil attack happens at a time. Initial ranks of the attacker A (before the attack) are varied from the top 1% to 95%. A will create a number of Sybils s (as large as 5% of the number of nodes). We will discuss here the results for a simple Sybil strategy, in which A has links to and form its Sybils. The Sybils are not connected to each other.

For the implementation of the CPR algorithm, MCL is chosen to derive clusters from the graph. The density metric is $r(i) = \frac{\#\ external\ Links\ of\ C_i}{2 \cdot (\#internal\ Links\ of\ C_i)}$. The mapping function is $f(i) = \epsilon_i = 0.15^{r(i)}$.

Figure 1 and 2 show the resilience of CPR against Sybil attacks, in comparison with the original PageRank (PR) metric. One striking fact is that under PR, attackers can always obtain the highest rank. We observed that the distribution of PR values in P2P (undirected) graphs has small variance, thus a small increase in PR value can result in great improvement in rank. Our reputation metric demonstrates substantial resilience against the attacks. In small-world graphs, the new ranks are very close to the initial ones and sometimes even higher (attackers get punished). In power-law graphs, attacks caused by very highly-ranked nodes are hard to detect, but as the initial rank of the attacker decreases, the metric gets more and more effective in detecting the attack. We also vary number of Sybil per

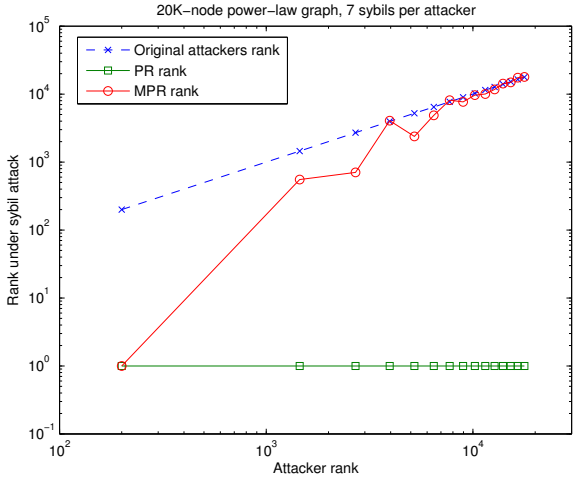

Figure 1. Resilience under power-law graph, 7 Sybils per attacker

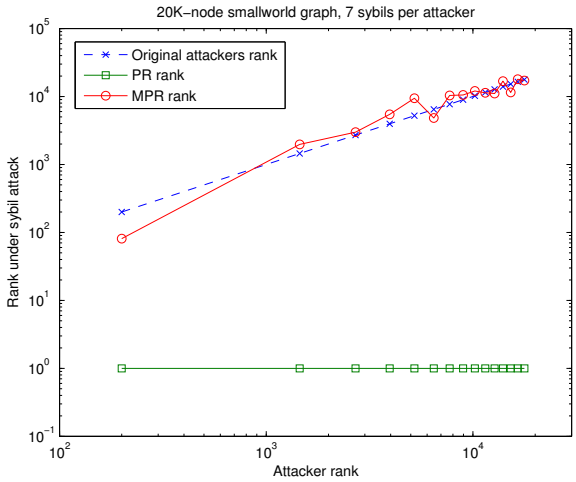

Figure 2. Resilience under small-world graph, 7 Sybils per attacker

attacker s (the result is not shown due to space constraint). We find that ratios $\frac{new\ rank}{old\ rank}$ vary around 1, even when the attacker controls 1000 Sybils (or 5% number of nodes in the network). Thus, CPR works well against powerful attackers.

To assess the *intuitiveness* of rankings produced by the CPR metric, in the absence of Sybil attacks, we use *normalized minimum Kendall distance* [3] metric to measure the similarity between two rank lists (produced by PR and CPR). The larger the distance, the more dissimilar the two lists are. We also compute the Kendall-distance of the top T ranks (first T elements of the lists). The results reveal the difference in K-distance values between power-law and small-world graphs. In either case, the top 100-rankings produced by CPR are quite similar than those produced by PR. More specifically, the K-distance values are always smaller than 0.4. Such values for the complete lists are even smaller: less than 0.05 for small-world graphs and 0.2 for power-law graphs.

6. Discussion

We have showed the proof of concept of how cluster information could help to alleviate the effect of Sybil attacks. In addition, there exists several open questions for the future work. First of all, scalability of the algorithm (in term of number of nodes it can handle) can be improved. Currently, MCL is the major bottleneck, as it requires a lot of large-matrix multiplications. On the other hand, our cluster-based PageRank model is generic and other graph clustering techniques can be investigated. Their performance (both in term of computation requirements and clusters quality) can be significantly different. Secondly, functions r and f in the CPR model can be chosen differently. r must generate meaningful values, i.e. dense clusters have the number of internal links considerably larger than the number of external links. Also, if r indicates a very dense region, f should give it a high value of ϵ. Thirdly, more formal, analytical study of the CPR algorithm is necessary. Finally and most importantly, it must be noted that we are studying our metric in the ideal model. As a result, we rely on a trusted, powerful party that collects data and constructs the complete reputation graph. For the metric to be used in real P2P environments, P2P protocols must be *realized* to securely implement this metric. Such realization process must at least constitute of realizing the distributed PageRank and distributed graph clustering computation.

7. Conclusion

In this paper, we have developed a Sybil-resilient reputation-based trust model for P2P systems. We proposed a new reputation metric based on the original PageRank metric and the graph clustering information. We employ the ideal/real model, which is mostly used in studying security protocols, to reason about the resilience of our metric against Sybil attacks. We showed that the PageRank metric is even easier to manipulate in P2P (or undirected) graphs than in Web (or directed) graphs. Our reputation metric (cluster-based PageRank or CPR) consists of a clustering algorithm and two functions r and f. Results from experiments with a particular implementation of CPR are encouraging. Specifically, Sybil attackers (even the powerful ones) are unlikely to be able to increase their reputation. In the absence of Sybil attacks, CPR produces meaningful and intuitive rankings, as they are similar to those produced by PageRank.

We identify several directions for the future work. Different clustering algorithms and choices of functions r and f could lead to very different performance. Formal analysis is necessary to yield more convincing arguments for our reputation metric. Finally, P2P protocols that implement PageRank and the graph clustering algorithms are needed so that our reputation metric can be fully realized in decentralized environments.

References

[1] C. Alice and E. Friedman. Manipulability of pagerank under sybil strategies. In *Proceedings of the first workshop of networked systems*, 2006.

[2] J. Douceur. The sybil attack. In *First international workshop on peer-to-peer systems*, pages 251–60, 2002.

[3] R. Fagin, R. Kumar, and D. Sivakumar. Comparing top k lists. *Journal on discrete mathematics*, 17(1):134–160, 2003.

[4] R. Levien. Advogato trust metric, 2004.

[5] M. Lik. *Computational models of trust and reputation: agents, evolutionary games, and social networks*. PhD thesis, MIT, 2002.

[6] S. Micali and P. Rogaway. Secure computation. In *Crypto 91*, 1992.

[7] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: bringing order to the web. Technical report, Stanford digital library technologies project, 1998.

[8] D. Stutzbach, R. Rejaie, and S. Subhabrata. Characterizing unstructured overlay topologies in modern p2p file-sharing systems. In *Proceedings of the Internet measurement conference 2005*, 2005.

[9] S. van Dongen. *Graph clustering by flow simulation*. PhD thesis, University of Utrecht, May 2000.

[10] R. Zhou and K. Hwang. Powertrust: a robust and scalable reputation system for trusted peer-to-peer computing. *IEEE transaction on parallel and distributed systems*, 18(4):460–473, 2007.