A comparison of two trust metrics

Jesse Ruderman
UCSD CSE 202 Fall 2004
December 15, 2004

Abstract

This paper describes and compares two trust metrics, Advogato and PageRank. It describes a novel attack against Advogato and gives an attack-resistance proof for PageRank.

1 Introduction

Trust metrics attempt to automate word-of-mouth reputation. A trust metric uses local information in the form of which users trust which other users in order to compute a global "trustworthiness" value for each user. Trustworthiness is computed relative to a "seed" of nodes initially assumed to be trustworthy.

The input to a trust is a directional graph. Each node is a user account and each edge represents a statement of the form "I, user A, certify that user B is trustworthy". A trust metric computes a number for each node, representing how much the system or trust seed should trust each node.

Advogato's attack model partitions nodes into three types:

- *Good* nodes behave well and have not certified any bad nodes.
- *Confused* nodes behave well but have issued certifications to bad nodes.
- *Bad* nodes are under the control of the attacker. The attacker controls who they certify and what they do.

One popular trust metric is the one used by open-source developer community website Advogato. The Advogato metric is based on network flow and decides which members appear to be trustworthy, competent open-source developers.

Another trust metric is PageRank, developed by the founders of Google. It is based on eigenvalues or random walks.

The rest of this paper is organized as follows. Section 2 describes several ways trust metrics are currently used. Section 3 attempts to define attack-resistance. Sections 4 and 5 describe Advogato and PageRank, respectively,

and section 6 compares these two trust metrics. Section 7 discusses some non-technical difficulties faced by systems that use trust metrics.

2 Uses for trust metrics

Advogato uses its trust metric in several ways. First, it makes the result for each user account public, helping open-source project owners decide how much to trust contributed code [6]. Second, it prevents uncertified members from doing certain things, such as posting articles to the Advogato front page.

The Google search engine treats each web page as a node and each link as a certification. It uses the PageRank metric to estimate the relative overall importance of web pages. Google uses this estimate of overall importance, combined with other factors, to rank search results in an attempt to show the most useful pages first.

Other proposed uses for trust metrics include protecting e-mail from spam [3] and protecting P2P networks [6].

3 Attack-resistance

While there is no agreed-upon definition of what it means for a trust metric to be *attack-resistant*, we can think of it as meaning that the attacker's success is bounded in some way by the number of confused nodes or some other property of the confused nodes. An attacker cannot defeat an attack-resistant trust network simply by creating a huge number of pseudonyms and connecting them in the right way; he must cause existing users to become "confused" and certify his nodes.

An attack-resistant trust network must have a trusted seed. Without one, an attacker could create a pseudonym corresponding to each existing user, and set up an identical certificaiton graph so the trust network cannot tell the difference.

An example of a trust metric that is attack-resistant but not useful is one that simply returns the seed. It is attack-resistant because bad nodes are not trusted at all, but it is not useful because it puts all of the burden of deciding trust on the choice of the seed.

An example of a trust metric that is not attack-resistant is one that accept all nodes within distance k from the seed, where $k \leq 2$ [3]. It is not attack-resistant because once an attacker convinces a "confused" node with distance $k-2$ or less to certify a bad node, the bad node can certify as many

bad nodes as it wants, and those nodes will all be within distance k from the seed.

4 Advogato's trust metric

Advogato's trust metric [1,7] is based on network flow.

First, the Advogato metric assigns a "capacity" c_x to each node x as a nonincreasing function of the distance from the seed. For example, advogato.org uses a capacity of 800 for the seed, 200 for the next two levels, 50 for nodes with distance 3 from the seed, and so on.

Each node A is then split into two parts $A-$ and $A+$, with a capacity-1 edge from $A-$ to the sink and a capacity-$(c_x - 1)$ edge from $A-$ to $A+$. A's certification of B becomes an infinite-capacity edge [7] from $A+$ to $B-$.

Advogato uses the Ford-Fulkerson algorithm to find the maximum flow. Since Ford-Fulkerson always picks the shortest augmenting path from the seed, any node with flow from $x-$ to $x+$ also has flow from $x-$ to the sink. Ford-Fulkerson takes $O(|f * ||E|)$, where $f*$ is the maximum flow. In this graph, $|f *|$ is simply the number of nodes accepted, so the algorithm takes $O(|V||E|)$.

Once network flow has been computed, the metric certifies each node for which there is flow from $x-$ to the sink. Since any node with flow from $x-$ to $x+$ also has flow from $x-$ to the sink, any node through which trust flows is itself certified.

4.1 Attack-resistance proof for Advogato's trust metric

Theorem: The number of bad serves chosen is bounded by $\sum_{x \in C}(c_x - 1)$, where C is the set of confused nodes.

Proof based on Levien and Aiken [1,7]: Consider the cut (S, T) where S consists of the good nodes (including the seed) and the compromised nodes, and T consists of the bad nodes along with the sink. The edges from S to T are edges from good and compromised nodes to the sink, and edges from compromised nodes to bad nodes. The edges from T to S are edges from bad nodes to good or confused nodes.

In any graph, the flow into the sink is equal to the flow across any cut. For this graph, this means that the total number of nodes chosen is equal to the flow across the cut. The flow across the cut is no greater than the sum of the flows along edges from S to T, or the sum of the number of good and confused nodes chosen with the flows along edges from confused nodes to bad nodes. Subtracting the number of good and confused nodes chosen from

both sides, we find that the number of bad nodes chosen is no greater than the flows along edges from confused nodes to bad nodes. The flow along such an edge is no greater than $c_x - 1$ because only chosen nodes can pass on flow. Thus the number of bad serves chosen is bounded by $\sum_{x \in C}(c_x - 1)$.

4.2 Problems with the Advogato metric

There are two major problems with Advogato's trust metric. First, the impact of a confused node "increases dramatically as it gets closer" to the seed [12]. There is little reason to believe that a node 4 hops from the seed is four times harder to trick than a node 5 hops from the seed.

The second problem exists even if we treat c_x as the "cost" to corrupt a node. This problem arises because the attacker can increase a confused node's c_x during the attack. Recall that c_x is assigned solely based on the distance from the seed to the node. An example of this attack follows.

In this example, nodes with some distance from the seed have capacity 400. Nodes with subsequent distances have capacities of 100, 25, 8, 2, and 1. Suppose an attacker confuses 400 nodes with $c_x = 1$ and a single node y with $c_y = 400$. This attack costs $\sum_{x \in C} c_x = 800$. Based on the proof above, one might expect the attacker to be unable to get more than 399 of his nodes trusted. But if one attacker node certified by y certifies the rest of the confused nodes, those confused nodes all become $c_x = 25$. The cost of the attack is only 800, but up to $\sum_{x \in C}(c_x - 1) = 400 \cdot 24 + 1 \cdot 399 = 9999$ bad nodes are accepted.

5 PageRank

PageRank [2] is a trust metric developed by the founders of the Google search engine. Instead of being based on network flow, it is based on eigenvectors or random walks.

PageRank has two definitions: a recursive formula, which is useful for understanding how PageRank is calculated, and a definition based on random walks, which is useful for understanding why PageRank is attack-resistant. This paper only presents the second definition of PageRank.

PageRank models the behavior of a "random surfer" in order to approximate the overall relative importance of web pages. The random surfer takes a random walk starting at a web page chosen using a seed probability distribution S. He then follows links randomly until he stops, stopping with probability p at each page he visits. The PageRank of a page is defined as the probability a random walk ends at the page.

PageRank can be tuned by changing the seed distribution S. One possible value for S is an even distribution over all web pages, but this S is not attack-resistant – an attacker can gain rank simply by creating new pages (and making sure Google's crawler finds them). Another possibility is to use a single page, such as http://dmoz.org/. While this works surprisingly well [2], it does not seem very fair. A third possibility is to distribute S evenly among the main page for each registered domain. This spreads out S almost as much as an even distribution, but it is harder to attack, because registering a domain costs about $10US.

PageRank can also be tuned by changing the stopping probability p. A smaller p makes PageRank more vulnerable to attack, while a larger p makes PageRank more sensitive to and more similar to S. p should be chosen to balance the need for attack-resistance with the need for the algorithm to make useful inferences given the trust seed and certificates (links).

An iterative calculation of Pagerank takes $O(|E|)$ per iteration. As described in the original PageRank paper, the Web has a property called "rapidly-mixing random walks" that allows PageRank to converge to for most Web pages in $O(\log |V|)$ iterations. Thus a useful approximation of PageRank only requires $O(|E| \log |V|)$ time to calculate.

5.1 Attack-resistance proof for PageRank

Let G be the graph before the attacker made any changes to it, and G' be the graph with the attacker's nodes and edges added. The attacker's edges include all edges from bad nodes, as well as all new edges from confused nodes.

Theorem: if S is 0 for all bad nodes, then pagerank of bad pages in G' is $\leq (\frac{1}{p} - 1)$ times the pagerank of confused pages in G.

Proof: The sum of the PR for the bad nodes is equal to the probabily that a random walk in G' ends in a bad node. Since no random walks start in bad nodes, the first bad node in a walk must be visited from a confused node. Thus the probability that a random walk in G' ends in a bad node is no greater than $(1 - p)$ times the probability that a random walk in G' ever hits a confused node. The probability that a random walk ever hits a confused node is the same for G and G'. Finally, the probabilty of ending at a confused node in G (equal to the total PR of confused pages in G) is no more than $\frac{1}{p}$ times the probability of hitting a confused node during a walk in G. In summary, the total PR for bad nodes is no greater than $((1 - p)\frac{1}{p}) = (\frac{1}{p} - 1)$ times the total PR for confused nodes.

Furthermore, the cost of confusing a node is roughly proportional to its

PageRank, since PageRank is a rough approximation of human attention. Sites with many visitors are more likely to protect themselves against vandals and hackers. They are also likely to charge more for link placement or for selling the entire site.

6 Comparison

Advogato's only advantage is that it returns only 1 or 0 for each node. Even this is usually considered a disadvantage, it is more useful than a real number in some situations. For example, it is more suitable for determining whether a user should be allowed to do something such as post an article to Advogato's front page.

PageRank has many advantages over Advogato. It is faster to calculate: PageRank can be approximated in $O(|E| \log |V|)$ while Advogato's metric takes $O(|E||V|)$ to compute. PageRank is more "fair" because two nodes with the same set of parents are guaranteed to have the same value. PageRank is more stable as nodes are added or removed. PageRank returns a real number for each node, which is suitable for ranking. While both Advogato and PageRank require a well-chosen "seed" of "initially trusted" nodes in order to be attack-resistant, PageRank is less sensitive to the precise seed chosen. PageRank's attack-resistance relies on a more reasonable cost function than Advogato's.

It is ironic that PageRank, which was designed to estimate human attention and interest paid to web pages, is more attack-resistance than Advogato, which was designed to be attack-resistant.

7 Non-technical considerations for trust metrics

Advogato's displaying of certifications creates social pressure. Some of the pressure is good: you are pressured to certify only trustworthy, competent open-source developers, lest people accuse you of being a weak point in the site's certification system. On the other hand, you may feel pressured to certify your friends as Masters, even though they really only deserve to be certified as Apprentices.

Advogato's certifications are also used for a diary-filtering feature. This feature uses a different trust metric and treats each reader as the seed when deciding which diaries the user should see. This encourages users to certify people who read interesting diaries rather than people who are trustworthy, competent open-source developers.

One problem with Advogato is that users often certify other users based on who those users claim to be, without verifying their identities. For example, one user signed up as "rms", claiming to be free software movement leader Richard Stallman. Even though the account is fake [8], 288 users certified the account as a Master level [9], putting it in fourth place for number of Master certifications [10]. Not surprisingly, the trust metric considers the "rms" account to be a Master.

One user even certified the account knowing that it was not Richard Stallman: "I certified it, even though I am reasonably sure it is not actually Richard M. Stallman. I did it because whether or not the account is real, RMS himself deserves props. I guess this creates the minor possibility that someone might try to use the account to impersonate RMS and say things he wouldn't say, but that has not happened yet." [11] It is not clear whether this user realized that the owner of the fake "rms" account could also cause damage by certifying other nodes that would misbehave.

Google is believed to discourage *bad* certifications by penalizing sites that link to known spam pages, or "bad neighborhoods" [5]. Since site owners like to get traffic from Google, this discourages sites from allowing their sites to link to spam sites.

References

1. R. Levien, A. Aiken. 2000. An Attack-resistant, Scalable Name Service. http://www.levien.com/fc.ps

2. L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: Bringing order to the web. 1998. http://citeseer.ist.psu.edu/page98pagerank.html

3. R. Levien. Attack Resistant Trust Metrics (draft thesis). July 2004. http://www.levien.com/thesis/compact.pdf

4. R. Levien. A few rating FAQ's. June 2002. http://www.advogato.org/person/raph/diary.html?start=225

5. Markus Sobek. PR0 - Google's PageRank 0 Penalty. http://pr.efactory.de/e-pr0.shtml

6. R. Dingledine, M. Freedman, D. Molnar. Draft of chapter 16, "Accountability", in "Peer-To-Peer: Harnessing the Power of Disruptive Technologies". http://freehaven.net/doc/oreilly/accountability-ch16.html

7. R. Levien. Advogato's Trust Metric. http://www.advogato.org/trust-metric.html

8. R. Dingledine. In irc://irc.freenode.net/p2p-hackers on 2004-12-13.

9. Page for account “rms” at Advogato. Certifications counted December 13, 2004. http://www.advogato.org/person/rms/

10. D. Stenberg. Advogato certificate stats. December 11, 2004. http://daniel.haxx.se/advogato/stats-2004-12-11.html

11. M. Stachowiak. Certifying the ‘rms’ account. July 2000. http://advogato.org/person/mjs/diary.html?start=67

12. R. Dingledine. In irc://irc.freenode.net/p2p-hackers on 2004-12-13.

13. J. Douceur. The Sybil Attack. 2002. http://citeseer.ist.psu.edu/douceur02sybil.html