Brief Announcement: Improving Social-Network-based Sybil-resilient Node Admission Control

Nguyen Tran Jinyang Li Lakshminarayanan Subramanian Sherman S.M. Chow
New York University, New York, NY, USA
{trandinh,jinyang,lakshmi,schow}@cs.nyu.edu

ABSTRACT

We present Gatekeeper, a decentralized protocol that performs Sybil-resilient node admission control based on a social network. Gatekeeper can admit most honest nodes while limiting the number of Sybils admitted per attack edge to $O(\log k)$, where k is the number of attack edges. Our result improves over SybilLimit [3] by a factor of $\log n$ in the face of $O(1)$ attack edges. Even when the number of attack edges reaches $O(n/\log n)$, Gatekeeper only admits $O(\log n)$ Sybils per attack edge, similar to that achieved by SybilLimit.

Categories and Subject Descriptors: C.2.4 [Computer Communication Networks]:Distributed Systems - distributed applications

General Terms: Algorithms, Design, Security

Keywords: Sybil attack, social network, Gatekeeper

1. INTRODUCTION

Open networked systems, such as Digg, YouTube, Facebook, BitTorrent, thrive on the participation of users in the form of content creation, sharing and voting. Unfortunately, these user-participation driven open systems are vulnerable to the Sybil attack where a single adversary can join the system using many bogus identities to amplify his attacking power. The root cause for the Sybil attack is the lack of a strong user identity thereby making it easy for an adversary to obtain new identities in the system. As a result, an adversary who launches the Sybil attack can use different identities to pollute the system with bogus information (in the form of content or votes) and affect the functioning of the system.

Social-network-based admission control offers the most promising defense against Sybil attacks. This approach seeks to limit the number of fake identities admitted into the system based on a social network. In particular, such a social-network-based Sybil defense exploits the assumption that an attacker has few social links to honest users since establishing these links often requires significant human effort. More concretely, the **Sybil-resilient Admission Control** problem can be stated as follows: Consider a social network G where each node represents a user and each link represents a trust relationship between two users. While most nodes in G are honest, some nodes are adversarial and an *attack edge* is a trust relationship between an honest and an adversarial node. Each user is initially aware of only their immediate neighbors in G and seeks to discover all the other honest users in G. An adversary can disrupt the discovery process of honest users by propagating Sybil identities across each attack edge. Sybil-resilient node admission control is a *decentralized* protocol that enables an honest user to discover and admit *most* honest users in the system while limiting the number of Sybil nodes admitted.

Social-network-based node admission control has been studied previously. SybilGuard [4] is the first work to show an admission protocol which limits the number of admitted Sybil identities to be $O(\sqrt{n}\log n)$ per attack edge, where n is the number of honest users in the social network. SybilLimit [3] significantly improves over SybilGuard and limits the number of Sybils admitted per attack edge to $O(\log n)$. In this brief announcement, we present a new protocol called Gatekeeper, that can limit the number of Sybils admitted per attack edge to $O(\log k)$, where k is the number of attack edges and the social network is random expander graph. Our result establishes optimality and improves over SybilLimit by a factor $\log n$ under the assumption of $O(1)$ attack edges. The security guarantee provided by Gatekeeper degrades gracefully with increasing numbers of attack edges; for example, in the face of $O(\log^C n)$ attack edges for any positive constant C, Gatekeeper admits $O(\log\log n)$ Sybil identities per attack edge. In the face of $o(n/\log n)$ attack edges, Gatekeeper achieves the same level of resilience as SybilLimit: both protocols admit $O(\log n)$ Sybils per attack edge with high probability. The proofs of these results and more details of the protocol can be found in [1].

To achieve these results, Gatekeeper uses an improved version of the *ticket distribution* process proposed in our prior work [2]. We have evaluated Gatekeeper on real-world social networks with varying number of attack edges. Our results show that Gatekeeper is able to drastically limit the number of admitted Sybil identities to a very small number while admitting almost all honest identities.

2. TICKET DISTRIBUTION

The principle building block of Gatekeeper is a ticket distribution protocol where each node acting as a ticket source disseminates n "tickets" throughout the social network. We originally designed the distribution algorithm for SumUp [2], a centralized Sybil-resilient vote collection system. SumUp performs max-flow computation from a vote collector to the set of voting users in order to limit the number of bogus votes cast by Sybil identities. It relies on ticket distribution to assign link capacities for the max-flow computation.

PODC'10, July 25–28, 2010, Zurich, Switzerland.
ACM 978-1-60558-888-9/10/07.

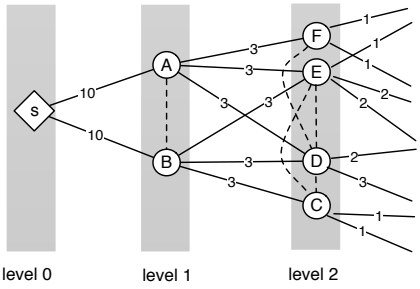

Figure 1: The ticket distribution process of a particular node S: The numbers on each link represent the number of tickets propagated via that link. The dotted lines are links between nodes at the same distant to the source

Gatekeeper uses ticket distribution differently. Our intuition is that, since the attacker only controls a small number of attack edges, each ticket source is relatively "far away" from most attack edges, resulting in few tickets propagated along an attack edge. Therefore, Gatekeeper can directly use a received ticket as a token for a node's admission.

We illustrate the ticket distribution process in Figure 1. Each ticket source disseminates $t = n$ tickets, where n is the number of honest nodes ideally. Tickets propagate in a "level-by-level" manner based on the shortest-path distance of a node to the ticket source. In Figure 1, source S attempts to disseminate 20 tickets. Each node is placed (conceptually) at a different level according to its shortest-path distance from S. S divides the tickets evenly and sends them to its neighbors. Each node keeps one ticket to itself and distributes the rest evenly among its neighbors at the next level. If a node does not have any outgoing links to the next level, it simply destroys all remaining tickets. The process continues until no tickets remain.

How to use ticket distribution for Sybil-resilient node admission? The naïve strategy for applying ticket distribution to admission control works as follows: each node admission controller (S) disseminates n tickets and accepts a suspect node if and only if it has received some tickets from S. Such a strategy has two inherent limitations. First, when distributing n tickets from one source, only a limited fraction of honest nodes ($\sim 60\%$ in simulations [2]) receive some tickets. Therefore, it will fail to admit a large fraction of honest nodes. Second, unlucky admission controllers that are relatively close to attack edges, will admit $\Theta(n)$ Sybil nodes. Our design of Gatekeeper addresses both limitations.

3. GATEKEEPER

The key idea of Gatekeeper is to perform distributed ticket distribution from *multiple* ticket sources. In the naïve strategy, an admission controller is also the ticket source. By contrast, in Gatekeeper, an admission controller explicitly picks m random nodes to act as ticket sources. Each randomly chosen ticket source distributes $t \approx c' \cdot n$ tickets, where c' is a constant derived in [1]. A node is called *reachable* from a ticket source if it consumes a ticket disseminated by the source. The admission controller admits a suspect node if and only if the node is reachable from at least $f \cdot m$ ticket sources, where f is a small constant such as 0.2.

Multi-source ticket distribution addresses both of limitations of the naïve strategy discussed earlier. Recall that the first limitation is concerned with the inability of admitting most honest nodes. Intuitively, Gatekeeper solves this problem because any honest node not reachable from one source can be reached by other sources. Ultimately, an honest node is admitted as long as it is reachable by $f \cdot m$ sources which is a high probability event. On the other hand, with a small number of attack edges, the attacker cannot appear close-by to many of the $f \cdot m$ sources, and thus is unlikely to receive a large number of tickets from as many as $f \cdot m$ sources to have many Sybils admitted. The second limitation is concerned with an unlucky admission controller (which is also the ticket source in the naïve strategy) close to some attack edges. Again, Gatekeeper solves this problem because the admission controller try to pick m random ticket sources as opposed to acting as the ticket source itself.

Below, we briefly discuss how Gatekeeper addresses the two remaining challenges in performing multi-source ticket distribution in practice: picking m ticket sources for each controller and estimating the number of tickets should be disseminated by each ticket source.

3.1 Choosing m random ticket sources.

To choose m random ticket sources, each controller node repeatedly performs random walks of length $l = O(\log n)$ from a randomly chosen neighbor to reach some ticket source. In a fast-mixing social network, a short random walk of length $O(\log n)$ reaches a destination drawn from the *node stationary distribution* of the graph. In other words, when the controller picks the end node of a random walk as a ticket source with probability $\frac{1}{d}$ (where d is the degree of that node), the resulting set of ticket sources are effectively chosen from all nodes at random.

3.2 Estimating the number of tickets needed.

Since a ticket source does not know n, it adaptively adjusts the number of tickets to be disseminated (t). A ticket source increases t if most of the current set of tickets manage to reach distinct nodes. In particular, under the assumption of an expander-like social graph, we can show that the number of nodes reachable by t tickets is at least some fraction $w = (e-1)/(c' \cdot 2e)$ of t. Therefore, if the ticket source observes that number of reachable nodes is greater than $w \cdot t$, it increases t and repeats the ticket distribution process. Otherwise, the ticket source terminates the adaption process with the current value t as a reasonable estimate for n. Such adaption is robust to manipulation by Sybil identities. Intuitively, since Gatekeeper limits the number of Sybil identities admitted per attack edge, the terminating condition occurs when t becomes greater than $c' \cdot n$ and is not affected by Sybil identities.

4. REFERENCES

[1] N. Tran, J. Li, L. Subramanian, and S. Chow. Optimal Sybil-resilient node admission control. *Technical report, http://www.news.cs.nyu.edu/~trandinh/GateKeeper_TR.pdf.*

[2] N. Tran, B. Min, J. Li, L. Subramanian. Sybil-resilient online content voting. In *NSDI*, 2009.

[3] H. Yu, P. Gibbons, M. Kaminsky, and F. Xiao. SybilLimit: A near-optimal social network defense against Sybil attacks. In *IEEE Symposium on Security and Privacy*, 2008.

[4] H. Yu, M. Kaminsky, P. Gibbons, and A. Flaxman. SybilGuard: defending against Sybil attacks via social networks. In *SIGCOMM*, 2006.