A Quick Survey of Social Network-based Sybil Defenses

Yazan Boshmaf
University of British Columbia, Vancouver, Canada

Abstract

Sybil (or multiple identity) attacks in distributed systems have serious security implications, as they can lead to service discontinuation and faulty business logic. Given that such attacks are inherently hard to solve, researchers have started to use information external to the system in order to develop new and effective defense schemes. Out of many proposals, social network-based Sybil defenses have had the greatest impact. In this short paper, we provide a brief survey of their design space and how different defense schemes compare to each other.

1 Introduction

In computer security, the *Sybil attack*, first coined by Douceur [9], represents the situation where a particular service in an identity-based system is subverted by forging identities. The security implications of the Sybil attack vary from out-voting honest users in online settings [11], to the corruption of routing tables in peer-to-peer systems [24].

This short paper presents a quick, compacted survey of effective Sybil defenses leveraging social networks. Since this approach of Sybil defenses via social networks was first introduced around seven years ago [28], it has attracted much more attention from the research community, rather than the mainstream industry.

We first provide the needed background and preliminaries in Section 2. After that, in Section 3, we start our journey through different proposals of social network-based Sybil defenses. Even though we put considerable effort into making the presentation self-contained, the curious reader can refer to the excellent surveys by Yu [26] and Viswanath et al. [25], which in effect highly influenced the content of this paper.

2 Background and Preliminaries

In what follows, we present background information and define the notations we use in the upcoming discussion.

2.1 Peer-to-Peer Systems

A *Peer-to-Peer* (P2P) system refers to a computer network in which each computer in the network can act as a client or server for the other computers in the network, allowing shared access to files and peripherals without the need for a central server [18]. In this paper, we focus on *Distributed Hash Tables* (DHTs): decentralized, distributed systems that provide a lookup service, similar to a hash table, for a group of collaborating peers or nodes. In a DHT, each node is assigned a unique identifier and is responsible for maintaining a set of (*key, value*) pairs of shared objects, which are stored and maintained in the DHT. Any participating node can efficiently retrieve the value associated with a given key of an object by following the DHT protocol (e.g., Chord [21], Kademlia [14]).

2.2 Social Networks

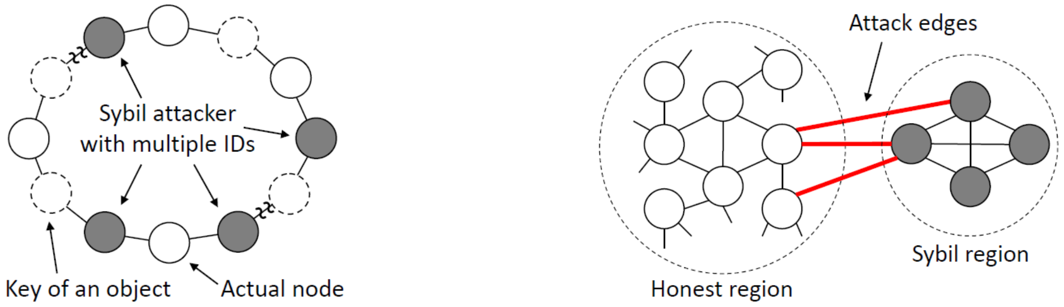
In mathematical sociology [12], a *social network* is defined as a social structure consisting of a set of actors (e.g., individuals, organizations) and a set of dyadic ties between these actors (e.g., relationships, connections, or interactions). Formally, a social network can be modeled as a graph $G = (V, E)$ where V represents a set of actors and E represents a set of social ties among these actors. A social tie between two actors can be either reciprocal such as friendship, or directional such as parenthood.

2.3 The Sybil Attack

The *Sybil attack* refers to the situation where an adversary controls a set of fake identities, each called a *Sybil*, and joins a targeted system multiple times under these Sybil identities [9]. The attack is named after the subject of the book Sybil by Flora Schreiber [20], a case study of a woman with multiple personality disorder who had 16 different "alters."

In this paper, we consider identity-based systems where each user is intended to have a single identity and is expected to use this identity when interacting with other users in the system. In such systems, we call a user with multiple identities a *Sybil user* and each identity the user uses a *Sybil identity*. In the Sybil attack, the adversary joins the targeted system using his Sybil identities and then mounts many follow-up attacks in order to disrupt the targeted system. For example, an adversary can pollute the voting scheme of a reputation system [11], subvert the routing and data replication services in DHTs [24], or cripple many critical functions of a wireless sensor network such as routing, resource allocation, an misbehavior detection [17].

Douceur [9] showed that without a centralized trusted party that certifies identities, Sybil attacks are always possible except under extreme and unrealistic assump-



(a) Distributed Hash Table

(b) Social Network

Figure 1: Social network-based Sybil Defenses (technical report of [5]). In (a), dashed nodes represent keys of objects drawn from an identifier space, and un-dashed nodes represent peers where each peer has a unique identifier drawn from the same identifier space. In (b), an edge is added between two nodes if these nodes trust each other, forming a network of trust among the peers in the DHT.

tions of resource parity and coordination among participating entities. Accordingly, traditional defenses against the Sybil attack rely on either trusting centralized party or bindings identities to resources that are hard to forge or obtain in large quantities, and as a result, preventing an adversary from creating many Sybil identities in the first place. For example, some approaches include solving memory or CPU-intensive cryptographic puzzles before granting access to system services [2, 4].

3 Social Network-Based Sybil Defenses

The goal of any *Sybil defense* is to prevent an adversary from gaining an advantage by creating and using Sybils. There are two classes of social network-based Sybil defenses: *Sybil detection*, which operates by detecting identities that are likely to be Sybils, and *Sybil tolerance*, which try to bound the leverage an adversary gains by using multiple Sybil identities. Both of these classes use the social network between the collaborating nodes (i.e., the peers) in the P2P system to defend against the Sybil attack, as depicted by Figure 1.

In this paper, we consider only social networks that are undirected and non-bipartite. In what follows, we treat each class of these defenses separately.

3.1 Sybil Detection

Many techniques [28, 27, 7, 23, 22] have been proposed that aim to label (i.e., detect) Sybils in the social graph by making the following three assumptions [25]:

1. The honest region is fast mixing [16], which means that it forms one tightly-knit community of users [10]. Put differently, random walks on the honest region can be modeled as an irreducible and aperiodic Markov chain [3]. This Markov chain is guaranteed to converge to a stationary distribution in which the landing probability on each node after sufficient steps is proportional to its degree.
2. The two regions are loosely connected, that is, the adversary cannot establish arbitrarily many social ties with non-Sybils. This means that there is a sparse cut between the honest and the Sybil regions such that there is a significant difference between their mixing times. The *mixing time* is defined as the maximum number of steps that a random walk needs to make so that the probability of landing at each node reaches the stationary distribution [3, 16]. It is assumed that the mixing time of the honest region is $O(\log n)$, where $n = |V|$ is the number of nodes in the graph.
3. The system is given one honest node in the graph.

Accordingly, one can find a sparse cut (i.e., partitioning) in the graph, and then declare the region where the known honest node belongs to as the honest region (see Figure 2 for an illustration). In the following, we give a brief summary of selected social network-based Sybil detection schemes.

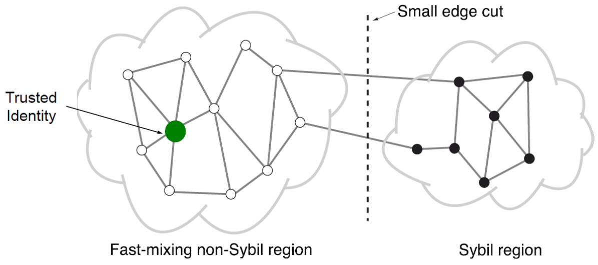


Figure 2: Sybil detection and general assumptions [25].

SybilGuard [28] and SybilLimit [27] are among the first Sybil detection schemes to be proposed. SybilGuard uses the intersections between modified random walks to determine whether identities should be given access to the system. SybilLimit improves on SybilGuard's bound by using multiple walks, which allows it to accept fewer Sybil identities per attack edge. Both of these schemes can be implemented in a centralized or decentralized fashion.

SybilInfer[7] is a centralized protocol that assumes full knowledge of the social graph. It uses a Bayesian inference technique that assigns to each node its probability of being Sybil. Unlike SybilGuard and SybilLimit, SybilInfer does not provide any theoretical bounds on the number of Sybil identities accepted per attack edge.

GateKeeper [23] is a decentralized Sybil detection protocol that improves over the guarantees provided by SybilLimit. It uses a variant of the ticket distribution algorithm used in SumUp [22] from multiple random identities in the graph to detect Sybils.

Even though social network-based Sybil detection schemes are relatively simple and easy to integrate into the system, they all suffer from inherent limitations. In particular, these schemes make strong assumptions about the topology of the social graph, where many of real-world social networks do not conform to these assumptions. For example, Leskovec et al. [13] show that real-world social networks have many small periphery clusters (i.e., tightly-knit group of nodes) that do not form one big community. Moreover, Mohaisen et al. [16] show that such social networks are generally not fast mixing. Moreover, extreme care should be taken in case the social network is imported from online social networking sites such as Facebook. Boshmaf et al. [5] showed that an adversary can perform automated social engineering to trick users into "befriending" his Sybils, and thus, there will be more attack edges than Sybils, rendering ineffective all detection schemes based on mixing times [26]. Consequently, these schemes have not found mainstream adaption, and they usually result in high false positive and false negative rates in real-world social networks, as depicted by Figure 4.

3.2 Sybil Tolerance

We now switch gears to Sybil tolerance schemes, which, unlike their Sybil detection counterparts, do not attempt to explicitly label identities as Sybil or non-Sybil. Instead, they aim to limit the leverage an adversary gains regardless to the number of Sybils he controls. In order to achieve this, these schemes use a *credit network* that is defined on top the social network between the nodes.

Credit networks [8, 6] were originally proposed in e-commerce to build transitive trust protocols in an environment where there are only pairwise trust accounts and no centralized trusted party. In such a network, identities (nodes) trust each other by offering pairwise credit (links) up to a certain limit. Nodes can use the credit to pay for services (transactions) they receive from each other. Accordingly, a transaction between two nodes is possible only if there is a path between them that has enough credit to satisfy the operation. Sybil tolerance schemes utilize credit networks and make the following two assumptions:

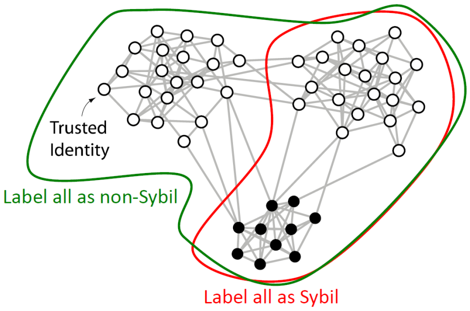


Figure 3: Sybil detection in social networks with multiple local clusters [25]. Given the trusted identity, the Sybil detection scheme can either accept all identities (high false negatives, in green), or reject access to the system for two clusters (high false positives, in red).

1. As with Sybil detection, it is assumed that the Sybil and the honest regions are loosely connected, that is, the adversary cannot establish arbitrarily many social ties with non-Sybils. Consequently, the same topological and stochastic properties discussed in Section 3.1 are also valid under the Sybil tolerance schemes (i.e., the cut size and the mixing time).
2. There exists a tractable strategy to (sub-)optimally assign credits to nodes in the network such that there is always liquidity in the network and all legitimate transactions are permitted. At the same time, this credit assignment limits illegitimate transactions to a bounded number of operations, which is proportional to the assigned credits along the attack edges (i.e., the edges of the sparse cut).

Accordingly, one can find such a credit assignment strategy in the graph, and then enforce the credit transaction scheme to allow a bounded illegitimate number of operations in the system, and thus, tolerating Sybils as shown in Figure 4. In the following, we give a brief summary of selected social network-based Sybil tolerance schemes.

Ostra [15] limits spam sent by users who create Sybil accounts. Ostra uses a social network, with credit values assigned to links. When a message is sent, Ostra searches for a path with available credits from the sender to the receiver. If no such path exists, the message is blocked. If a path does exist, the credit is transferred from each user to the next along the path.

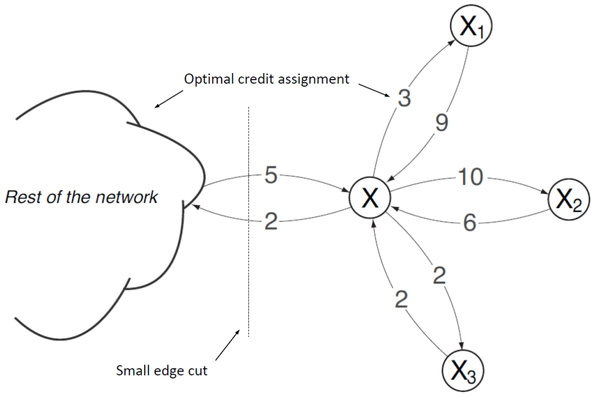


Figure 4: Sybil tolerance and general assumptions [25]. In this credit network, there are four Sybil nodes $\{X, X_1, X_2, X_3\}$. A directed edges (X_i, X_j) represents how much credit is available to X_i from X_j. For example, X can perform at most three consecutive transactions with X_1 (one way), after which the credit on the edge (X, X_1) is zero and on the edge (X_1, X) is 12.

Bazaar [19] protects buyers and sellers in online marketplaces like eBay by limiting the reputation manipulation (i.e., collusion) that is possible through the creation of Sybil accounts. It uses an expensive max flow-based techniques to estimate the reputation of users involved in a transaction and then flags fraudulent transactions.

SumUp [22] secures online voting against users who create Sybil accounts and vote multiple times. SumUp chooses a vote collector in the network and distributes tokens (i.e., credits) on the links in the network inside a voting "envelope." Voters must find a path to the vote collector with available credits in order to cast a vote.

Social network-based Sybil tolerance schemes leverage both the social network and the transaction history to limit illegitimate operations, which results in high classification accuracy when compared to Sybil detection. Moreover, these schemes work on labeling transactions, and thus, only graceful degradation is expected in the presence of errors. Still, these schemes are limited to applications for which appropriate system properties are known to suite Sybil tolerance. In addition, it expected that deploying a "good" credit network to be challenging, where liquidity is preserved and illegitimate transactions are limited. Otherwise, the system can be susceptible to a Denial of Service (DoS) attack on the credit network, as illustrated in Figure 5.

4 Open Problems

To date, there is no comprehensive evaluation of how effective social network-based Sybil defenses are when the underlying assumption are completely or partially false.

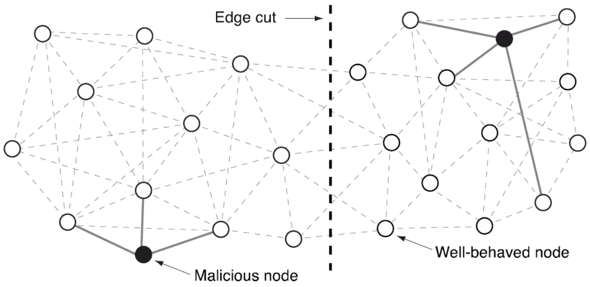


Figure 5: DoS in Sybil tolerance schemes [25]. Consider the scenario where the malicious node has more credit that the well-behaved node. The malicious node can now start draining the credits across the cut until it successfully denies the well-behaved nodes from executing transactions with nodes located in the left partition.

In particular, the assumption of the limited social capability of an adversary is quickly becoming a major concern, which is the focus of the following research questions:

1. How can we develop an approximate model that simulates (i.e., predicts) how Sybils "befriend" users under different assumptions of the adversary capabilities?
2. How can we use this model to evaluate the effectiveness of existing Sybil detection and tolerance algorithms?
3. What are the insights from such an evaluation that can be used to design Sybil defenses that resilient to even "social" adversaries?

5 Acknowledgments

We would like to thank our colleagues at LERSSE and NetSysLab for their continuous support and encouragement. Moreover, we would like to thank the folks at ICICS Reading Room for the unlimited supply of miniature Easter creme eggs (the Goo is back! [1]).

References

[1] Cadbury creme egg on facebook. `https://www.facebook.com/cadburycremeegg`, 2012.

[2] M. Abadi, M. Burrows, M. Manasse, and T. Wobber. Moderately hard, memory-bound functions. *ACM Transactions on Internet Technology (TOIT)*, 5(2):299–327, 2005.

[3] E. Behrends. *Introduction to Markov chains*, volume 228. Vieweg, 2000.

[4] N. Borisov. Computational puzzles as sybil defenses. In *Peer-to-Peer Computing, 2006. P2P 2006. Sixth IEEE International Conference on*, pages 171–176. IEEE, 2006.

[5] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu. The socialbot network: when bots socialize for fame and money. In *Proceedings of the 27th Annual Computer Security Applications Conference*, ACSAC '11, pages 93–102, New York, NY, USA, 2011. ACM.

[6] P. Dandekar, A. Goel, R. Govindan, and I. Post. Liquidity in credit networks: A little trust goes a long way. *Arxiv preprint arXiv:1007.0515*, 2010.

[7] G. Danezis and P. Mittal. SybilInfer: Detecting sybil nodes using social networks. In *NDSS*, Feb. 2009.

[8] D. DeFigueiredo and E. Barr. Trustdavis: a non-exploitable online reputation system. In *E-Commerce Technology, 2005. CEC 2005. Seventh IEEE International Conference on*, pages 274–283, July 2005.

[9] J. R. Douceur. The sybil attack. In *IPTPS '01: Revised Papers from the First International Workshop on Peer-to-Peer Systems*, pages 251–260, London, UK, 2002. Springer-Verlag.

[10] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75 – 174, 2010.

[11] A. Jøsang and J. Golbeck. Challenges for robust of trust and reputation systems, Sept. 2009.

[12] D. Knoke and S. Yang. *Network Analysis (Quantitative Applications in the Social Sciences)*. Sage Publications Inc., 2007.

[13] J. Leskovec, K. Lang, A. Dasgupta, and M. Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1):29–123, 2009.

[14] P. Maymounkov and D. Maziéres. Kademlia: A peer-to-peer information system based on the xor metric. In *Revised Papers from the First International Workshop on Peer-to-Peer Systems*, IPTPS'01, pages 53–65, London, UK, 2002. Springer-Verlag.

[15] A. Mislove, A. Post, P. Druschel, and K. Gummadi. Ostra: Leveraging trust to thwart unwanted communication. In *Proc. of NSDI*, 2008.

[16] A. Mohaisen, A. Yun, and Y. Kim. Measuring the mixing time of social graphs. In *Proceedings of the 10th annual conference on Internet measurement*, pages 383–389. ACM, 2010.

[17] J. Newsome, E. Shi, D. Song, and A. Perrig. The sybil attack in sensor networks: analysis defenses. In *Information Processing in Sensor Networks, 2004. IPSN 2004. Third International Symposium on*, pages 259 – 268, april 2004.

[18] A. Oram. *Peer-to-peer: Harnessing the power of disruptive technologies*. O'Reilly & Associates, Inc., 2001.

[19] A. Post, V. Shah, and A. Mislove. Bazaar: strengthening user reputations in online marketplaces. In *Proceedings of the 8th USENIX conference on Networked systems design and implementation*, NSDI'11, pages 14–14, Berkeley, CA, USA, 2011. USENIX Association.

[20] F. Schreiber, S. Martínez, and L. Vigil. *Sybil*. 1981.

[21] I. Stoica, R. Morris, D. Liben-Nowell, D. R. Karger, M. F. Kaashoek, F. Dabek, and H. Balakrishnan. Chord: a scalable peer-to-peer lookup protocol for internet applications. *IEEE/ACM Transactions on Networking (TON)*, 11(1):17–32, 2003.

[22] N. Tran, J. Li, L. Subramanian, and S. Chow. Optimal sybil-resilient node admission control. In *INFOCOM, 2011 Proceedings IEEE*, pages 3218 – 3226, april 2011.

[23] N. Tran, B. Min, J. Li, and L. Subramanian. Sybil-resilient online content voting. In *Proceedings of the 6th USENIX symposium on Networked systems design and implementation*, NSDI'09, pages 15–28, Berkeley, CA, USA, 2009. USENIX Association.

[24] G. Urdaneta, G. Pierre, and M. van Steen. A survey of DHT security techniques. *ACM Computing Surveys*, 43(2), Jan. 2011. `http://www.globule.org/publi/SDST_acmcs2009.html`.

[25] B. Viswanath, M. Mondal, A. Clement, P. Druschel, K. P. Gummadi, A. Mislove, and A. Post. Exploring the design space of social network-based sybil defenses. In *Communication Systems and Networks (COMSNETS), 2012 Fourth International Conference on*, pages 1–8, January 2012.

[26] H. Yu. Sybil defenses via social networks: a tutorial and survey. *SIGACT News*, 42:80–101, October 2011.

[27] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao. Sybillimit: A near-optimal social network defense against sybil attacks. In *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, pages 3–17, Washington, DC, USA, 2008. IEEE Computer Society.

[28] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman. Sybilguard: defending against sybil attacks via social networks. *SIGCOMM Comput. Commun. Rev.*, 36:267–278, August 2006.