A Survey on Sybil Attacks and Defenses

Xinghuang Xu
EECS Department
Wichita State University
Email: xxxu3@wichita.edu

***Abstract*—Sybil attack has been a thread to most peer to peer network systems. If new identities can be created without control in a distributed system, the system is susceptible to Sybil attacks. For example in IMDB, sybil accounts can be created to boost the score of a new movie in order to attract potential audiences to watch the movie in theatres. The survey paper is a guideline for open distributed system designers who want to introduce defense mechanisms into their systems to protect against Sybil attacks. We first define various Sybil attacks under different domains with different goals then we presents three categories of defenses against Sybil attacks. The three categories include the traditional approach, the social network based approach and the domain specific approach. We also analyse the differences among the three categories and the advantage/disadvantage of methods within each category. Readers will have a deeper understanding of how to protect their distributed systems against Sybil attacks after reading this survey.**

I. Introduction

Sybil is book written by Flora Rheta Schreiber about the treatment of Sybil Dorsett for dissociative identity disorder. She is believed to have manifested sixteen different personalities according her doctor Cornelia B Wilbur.[1]

This survey is about a specific system security attack name Sybil Attack. Sybil attack takes place when an adversary in a system acts as if he is multiple users with different identities in order to disrupt the proper functionality of the underlying system and benefit himself. Sybil attack has proven to be harmful in many systems. For example, in a book recommender system. The popularity of a book depends on the number of people who have liked it. In such a system, the goal is to find books that is likely to be of interest to users based on others' recommendations. An attacker can create many fake accounts and out vote legitimate users on the books he/she wants to promote or demote. The success of the attack is almost guaranteed given the fact that most legitimate users are too lazy to vote in a recommender system. There are many other domains that are vulnerable to Sybil attacks. It's impossible to enumerate all the domains susceptible to Sybil attack in this survey, so we have selected some typical domains that are well studied in section II. If your domain is not listed in II, that doesn't mean your domain is free from Sybil attacks. You should ask yourself if it's possible to create fake identities in your system, if your answer is yes then you should keep an eye open for Sybil attacks.

The following is a roadmap of the rest of the paper. In section II, we define some common properties of sybil and a list of some major domains that are susceptible to Sybil attacks are provided. We define the goal of the defense and three main types of defense mechanisms in section III. The three categories of Sybil defenses are traditional defense, social network based defense and domain specific defense. We provide descriptions on how the defense works and point our their advantage/disadvantage in terms of their efficiency, false positive/negative rate, deploy-ability and more. In the last section **??**, we would conclude this survey. After reading this survey you will have a deeper understanding of Sybil attacks and should be equipped with many techniques to defense against Sybil attacks.

II. Sybil Attacks

A. Attack Model

The context of the attack can be vastly different. Sybil attacks can take place in vastly different domains such as Wireless Sensor Network, Online Social Network, Reputation System, Ad hoc Mobile Network and etc. The goal of the attacker can vary too. The goal of an attacker can either be to control the system for self benefit or to subvert the normal functionality of the system. Attackers with the goal to manipulate the system will create sybil nodes that camouflage themselves as honest nodes and acts like honest nodes. For example, in an Online Social Network, the attacker can gradually create sybils and make them looks like real users by using other people's online profile and daily posts etc. Some attackers would go further to ensure their fake accounts act like regular users with regular logins, friend requests, friend request acceptances and real user like click streams. After the sybils have made enough connection with honest users, they are start spreading news or malware to disrupt the targeted online social network. If the goal of the attacker is to subvert the targeted system, he/she will usually inject as many bad players into the system as possible. Sybils are like bombs that are being hide in the system and when they explode at once, the system could be destroyed.

In sum, attackers can create sybils quickly or gradually. Sybils can have a short or long camouflage periods when they act as honest players. Sybils can launch the attack all at once or they can misbehave one at a time. Sybils can have no connection within themselves or they can form relationship between each other to form a group.

B. Attacks in Different Domains

1) Routing: In a distributed system, routing required the participation of many nodes. For example, when node A wants to look up node D, it will ask its neighbors and its neighbors

would go ask their neighbors and on and on until someone has the knowledge of D and pass Ds location all the way back to A through a path. An attacker can inject many malicious nodes inside and network to disrupt the routing. In a sybil infected network, when A ask about Ds location, if the lookup path pass a malicious node, the malicious node can return a false address or can do nothing. Either way, the process of lookup will be slow down greatly. With sufficient proportions of sybil nodes, attackers can block communications among nodes altogether and render the system useless.

There are concrete real world examples of Sybil attacks on routing protocols. Geogaphical routing protocol is vulnerable to Sybil attacks because it requires nodes to exchange coordinate data with their neighbors to efficiently address packets. By using the Sybil attack, an attacker can create multiple identities in different geographical locations thus making him available in multiple places at once which violates the fundamental assumption of the routing protocol[2]. Sybil attack pose a thread to the seemingly robust multipath routing protocol too.[2]

The popular Distributed Hash Tables (DHT) which underlies many peer-to-peer systems are also known to be vulnerable to Sybil attack. In DHT that are opened to the rest of the world like Vuze DHT, an adversary can introduce a large number of corrupt nodes in the network to degrade of the performance of the targeted DHT.[3]

2) Content Rating System: Sybil attack is a fundamental thread to any user-based content rating system such as Goodreads, Youtube and IMDB. There are hugh incentives in this kind of attacks because attackers can promote low-quality content to a wide audience. For example, it has been studied that many people check the IMDB score before going to see the movie in theater. A high IMDB score will attrack more audiences to go to the theater thus making the movie more profitable. This is not hypertheticcal. There are successful real world cases. For example, the famous Slashdot poll on the best compute science school has caused students to write automatic scripts to vote for their schools repeatedly. Moreover, some underground companies made money through assisting clients in promoting their Youtube video's view counts by using a large number of Sybil accounts.[4]

3) Reputation System: Sybil attack post a significant challenge for building repuations systems. In a repuation system, an adversary can create a large number of identities and maliciously increase the reputation of one or more master identities by giving false recommendations to them. Sybils can also promote their own reputations and falsely accuse well-behaved players in the system to hurt their reputation. For example, in eBay.com repuation is calculated as the sum of (+1,0.-1) of all the trasaction ratings no matter how big the transaction is. Sybils can be create to make small transactions with a seller and automattically give them good reviews to boost their reputation. Afterward, the seller can use that repuation on a dishonest transaction of high value. By using Sybil attack, a dishonest seller can hide the fact he frequently misbehaves at a certain rate.

Moreover, in networks that use repuation scheme to find misbehaving nodes/sybils, nodes with good reputation can report nodes they believe to be misbehaving in its neighbors. But this scheme can backfire. For example, users can collude to artificially boost the reputation values of one or more friends, or falsely accuse well-behaved users of misbehavior. when adversaries control enough nodes and decide to repeatedly report honest nodes. The outcome is that most of the honest nodes will be considered malicious and be removed from the networks, the malicious nodes afterward will take full control of the whole system and use it for their their own benefits. Detecting such collusion attacks is yet an unsolved problem that severely limits the impact of existing reputation systems.[5][6]

4) Peer-to-Peer System: Sybil attack can be used to gain a disproportional share of resources in P2P network.

In a distributed system, people are sharing their resources such as bandwidth, memory and data. An adversary can create sybils to claim an unfair and disproportionate share of the resources.

Also in distributed storage, sybils can cause data lost by being selfish and not storing the fragment of data that are asked to store. Sybils can be used to degrade the performance of the distributed file system by not responding to file request or provide the wrong file segment. What's more, because some file systems replicate data to neighbor nodes, sybils can be used to crawl the entire file system through frequently hopping into different areas in the network.[6][7]

III. Sybil Defense

The problem of defending against sybil attacks has been thoroughly study but there are no good known method that could completely eliminate the problem. Some of the centralized solution claim to be able to eliminate Sybil attacks at the price of adding authentication overhead to the system or sacrifising the open nature of distributed system. Most of the approaches study in our survey are seeking to reduce the effect of Sybil attacks in their systems instead of eliminating them. There is always a trade of between efficiency, security and system compelxity in all the approaches. In this section, we first classify the defense based on its timeline and then on the underlying technique they use. We include a short description for each teachnique which will be extended in the final paper.

A. Traditional Appraoch

1) Trusted Certification: This is the most popular solution for countering Sybil attacks, it required a trusted certifying authority that validates the identity of a node before it joins the system. There are two variations in this approach. One is the centralized version, the other is the semi-centralized version. In the centralized version, it is assumped that there is a trusted central authority who can verify the validity of each participant. After the validation, a certificate will be given to each participant. The participant then can use the certificate to access the system. The model is very popular and has been used widely. Most authentication services use this kind of

model. The semi-centralized approach seek to cut of the cost of asymmetric criptography used in the centralized version. It leverage a techniqe called partical identity verifications. The approach still need to rely on a trusted base station but reduce the involvement of a third party authority.

The problem of trusted certification approach is that it rely on a centralized trusted authority for credential generation, assignment and verification. However, it sacrifice the open nature that underlies the success of these distributed systems and increase the overhead of the system. [8][9][10].

2) Resource Testing: Resource Testing is another line of solution. The idea behind resource testing is that each identity should own a fair amount of resource because it runs on a legitimate cilent otherwise there is a high potential that this is a sybil node. The question is how can we test that there are resource backing up a node? Some propose the testing of IP address because multiple identities sharing a single IP address is a good sign of Sybil attacks. Others test resources such as computing power, network bandwidth. A variation of the resource testing method is called Recurring Costs. For example, in one solution participants are required to perform some tasks such as solving puzzles[11] periodically. Turing tests are also suggested as a recurring cost solution[12]. With Recurring Costs, the cost of Sybil attacks have become more expensive but would the benefit still outweight the cost? Cloud services have definitely help drive down the cost of Sybil attacks.

B. Domain Specific Approaches

1) Ad hoc Networks: In wireless ad hoc networks, a group of sybils are usually sharing the same device and they can be detected through monitoring signals features or the moving patterns of coexisting identities. SybilCast is a novoel protocol proposed in [13] that can limit the number of fake identities in centralized multichannel wireless networks. SybilCast can ensure that each honest user gets at least a constant fraction of their fair share of the bandwidth and complete his or her data download in asymptotically optimal time.

2) Wireless Sensor Netowrk: In wireless sensor netowrk, Demirbas et al. proposed an sybil attack counter measurement by using received signal strength indicator (RSSI). The algorithm proposed in [14] claims to be light weight because it only require the colloboration of one other node apart from the receiver and accurate because it detects sybil attack cases with 100% completeness and only a few percent false positives. [14]

C. Social Network-Based

Yu et al. has started a new era of sybil defense when he proposed the idea of detecting sybils using a unique structure in the social network graph. Even though attackers can inject many sybils into a social graph, the connections between honest users and sybils are limited[15]. For example, honest users on facebook would not randomly accept friends if they do not know the person. Suprisingly, the social network appraoch has showed to be able to overcome some of the earlier approaches limitations and shortcomings.

1) SybilGuard: SybilGuard designed by Yu et al. [15] is one of the first Sybil defense techniques based on Social Network. The approach assumes that each edge in the graph between two identities indicates a human-established trust relationship and malicious users can only create limited edges between honest users. SybilGuard bounds the number of malicious sybils a user can create by exploiting the property that there exist a disproportionaltely small "cut" in the graph between the sybil nodes and the honest nodes.

2) SybilLimit: The approach take by SybilLimit in [16] is the same as SybilGuard but SybilGuard can dramatically reduce the number of sybil nodes accepted by a factor of $\sqrt{n}$.

3) SybilInfer: SybilInfer takes the approach of labelling nodes in a social network as honest users or Sybils. Internally, it uses a probabilistic model of honest social networks as its knowledge base and an inference engine to obtain the potential regions of dishonest nodes. It claims to be more accurate and more applicable when compare to both SybilGuard and SybilLimit.

IV. Project Deliverable

We have listed potential threads of Sybil attacks under different context and showed different types of counter measurements. We want the final paper to serve as introduction and guidline for sybil attacks and defenses. To obtain this goal, we would like to present more detail explaintion of each Sybil defense approaches in the final survey. Furthermore, a comphrehensive comparison among the different defense mechanisms will be included. If time permits, we would also look into evaluating some of the defense mechanisms by using metrics like false negative/positive rate, the time and code complexity the soution would add to the existing system. Also we would like to point out potential directions/opportunities in the research of sybil attacks/defenses in the final survey paper.

References

[1] Wikipedia, "Sybil (book)," 1973. [Online; accessed 02-May-2016].

[2] C. Karlof and D. Wagner, "Secure routing in wireless sensor networks: Attacks and countermeasures," 2003.

[3] G. Danezis, C. Lesniewski-laas, M. F. Kaashoek, and R. Anderson, "Sybil-resistant dht routing," in *In ESORICS*, pp. 305–318, Springer, 2005.

[4] 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, (Berkeley, CA, USA), pp. 15–28, USENIX Association, 2009.

[5] G. Swamynathan, K. C. Almeroth, and B. Y. Zhao, "The design of a reliable reputation system," vol. 10, pp. 239–270, Dec. 2010.

[6] Q. Lian, Z. Zhang, M. Yang, B. Y. Zhao, Y. Dai, and X. Li, "An empirical study of collusion behavior in the maze p2p file-sharing system," in *In ICDCS*, 2007.

[7] S. Wolchok, O. S. Hofmann, N. Heninger, E. W. Felten, J. A. Halderman, C. J. Rossbach, B. Waters, and E. Witchel, "Defeating vanish with low-cost sybil attacks against large dhts," 2009.

[8] 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*, pp. 259–268, April 2004.

[9] M. Castro, P. Druschel, A. Ganesh, A. Rowstron, and D. S. Wallach, "Secure routing for structured peer-to-peer overlay networks," *SIGOPS Oper. Syst. Rev.*, vol. 36, pp. 299–314, Dec. 2002.

[10] A. Adya, W. J. Bolosky, M. Castro, G. Cermak, R. Chaiken, J. R. Douceur, J. Howell, J. R. Lorch, M. Theimer, and R. P. Wattenhofer, "Farsite: Federated, available, and reliable storage for an incompletely trusted environment," *SIGOPS Oper. Syst. Rev.*, vol. 36, pp. 1–14, Dec. 2002.

[11] N. Borisov, "Computational puzzles as sybil defenses," in *Proceedings of the Sixth IEEE International Conference on Peer-to-Peer Computing*, P2P '06, (Washington, DC, USA), pp. 171–176, IEEE Computer Society, 2006.

[12] L. V. Ahn, M. Blum, N. J. Hopper, and J. Langford, "Captcha: Using hard ai problems for security," in *Proceedings of the 22Nd International Conference on Theory and Applications of Cryptographic Techniques*, EUROCRYPT'03, (Berlin, Heidelberg), pp. 294–311, Springer-Verlag, 2003.

[13] C. Zheng and D. S. Gilbert, "Thwarting sybil attacks and malicious disruption in wireless networks, http://www.comp.nus.edu.sg."

[14] M. Demirbas and Y. Song, "An rssi-based scheme for sybil attack detection in wireless sensor networks," in *World of Wireless, Mobile and Multimedia Networks, 2006. WoWMoM 2006. International Symposium on a*, pp. 5 pp.–570, 2006.

[15] H. Yu, M. Kaminsky, P. B. Gibbons, and A. D. Flaxman, "Sybilguard: Defending against sybil attacks via social networks," *IEEE/ACM Transactions on Networking*, vol. 16, pp. 576–589, June 2008.

[16] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao, "Sybillimit: A near-optimal social network defense against sybil attacks," in *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pp. 3–17, May 2008.