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 III. If your domain is not listed in III, 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 III, 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 IV. 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. 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.

III. Sybil Attacks

A. Routing System

Routing is an essential part of all distributed systems. Many P2P systems use DHT to store routing tables. There are two major strategies to perform routing table attacks in such P2P

systems named horizontal attack and vertical attack according to [2]. Assume a setting of Kademlia-based system[3] where the system has N nodes, each node maintains a k-bucket routing table and the average number of hops in routing a message in the system is $O(log(N))$.

A horizontal attack aims at polluting as many routing tables as possible through spreading sybils widely across the whole system. Based on the k-bucket mechanism, an attacker need only to control at least one sybil among a node's neighbours to effectively intercept messages. To launch a successful horizontal attack, an attacker would need to roughly inject $\frac{N}{min(k,O(log(N)))}$ sybils into the system to hijack the whole system. The next question concern the attack is how much resource is needed to perform this attack. A straightforward way to do it is to run a sybil instance in one machine but this is highly inefficient. By modifying the DHT client, attackers can run many sybil instances simultaneously on a machine. Moreover, by exploiting the hopping technique in [4], sybils can change their id periodically and jump to a new location in the DHT. By jumping into new locations periodically, a sybil instance can cover a lot more area than a static sybil.

On the other hand, vertical attack tries to insert as many sybils as possible into one specific routing table. Using vertical attack, a specific content ID can be targeted. This attack would be made more difficult with a very large DHT and if the DHT protocol assign random ids to newly join nodes but this is not the case in Mainline DHT(MLDHT)[2]. MLDHT allows nodes to pick their own IDs and this security weakness has been around for over a decade and no one care enough to fix it. Given a target Id, sybils can generate ids that will locate them close to the target and 'isolate' the target from the others.

The authors in [2] also mention a hybrid approach in which attacks would first launch a horizontal attack to take control over the whole system then target individual nodes with vertical attack. In sum, a hybrid attack would lead to the attacker controlling the whole system.

There are other routing protocols that are vulnerable to Sybil attack. Geographical routing protocol requires nodes to exchange coordinate data with their neighbours to efficiently address packets. By using 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 geographical routing protocol[5]. Sybil attack poses a thread to the seemingly robust multipath routing protocol too. For more detail please see [5].

B. Content Rating System

Sybil attack is a fundamental thread to many user-based content rating system such as GoodReads, YouTube and IMDB. There are huge incentives in this kind of attacks because attackers can promote low-quality content to a wide audience. It has been studied that many people check the IMDB score before going to the movie theatre. A high IMDB score will attract more audiences thus making the movie more profitable. This is not hypothetical. There are successful real world cases. For example, the famous Slashdot poll on the best computer 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.[6]

C. Online Market Place

Sybil attack has posted a significant challenge for building reputation systems in online market place. In a reputation 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 reputation is calculated as the sum of (+1,0.-1) of all the transaction ratings no matter how big the transaction is. Sybils can be create to make small transactions with a seller and automatically give them good reviews to boost their reputation. Afterwards, the seller can use that reputation 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 reputation scheme to find misbehaving nodes/sybils, nodes with good reputation can report nodes they believe to be misbehaving in its neighbours. 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 misbehaviour. When adversaries control enough sybil 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 will take full control of the whole system and use it for their own benefits. Detecting such collusion attacks is yet an unsolved problem that severely limits the impact of existing reputation systems.[7][8]

D. Resource Sharing System

Sybil attack can be used to gain a disproportional share of resources in P2P network. In a P2P system, people can share their resources such as bandwidth, memory, computation power and file. An adversary can create sybils to claim an unfair and disproportionate share of the resources that were intended to be distributed amongst all nodes in the system. For example, in a public cloud infrastructure like Amazon Web Services, Google Cloud, each user is eligible for a free 15GB of data storage. An attacker can create 100 or more sybil accounts and claim more than 1500GB of free storage. Let's consider a distributed file sharing system, the download speed depends on how much credit a user. To obtain credits, a user needs to make contribution and upload files to others. This seems like a fair file sharing system until an attacker creates many sybil accounts and generate credits from download/upload files between themselves. With Sybil attack, the attacker can generate unlimited credits and use them to quickly download files from others.

E. Distributed Storage System

In distributed storage system, nodes are required to score a fragment of a file and each fragment is duplicated into multiple machines to prevent data lost and increase the performance of file download speed. Sybils can cause data lost by being selfish and not storing the fragment of data that they are asked to store. Sybils can also 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 distributed file systems replicate data to neighbouring nodes, sybils can be used to crawl the entire file system through frequently hopping to different locations in the network and obtain data fragments from all its neighbours.[8][4]

F. Online Social Network

Online Social Networks(OSN) like Facebook and Twitter are vulnerable to Sybil attack as well. One goal of the attacker can be to crawl the OSN's user's personal data. Personal data includes name, phone number, age, address etc. In order to obtain those personal information attackers need to be friends with the actual users to be able to see those information. An attacker in this case will create a lot of fake accounts that camouflage themselves as real users. New sybil accounts can be created by copying the information of some of the victims who befriended a sybil account and have their personal information stolen. New sybils will continue to send friend requests to other people and some will even befriend themselves. After using sybils to crawl users' account information, attackers can then make a profit from selling those information. Another goal of attacking the OSN can be to spread spam. More and more people these days use facebook and twitter as their news channel and read posts on the Social network as they read news paper. After successfully befriend many users and their friends, a sybil account can more effectively spread spam or even malicious files.

G. Collaborative Mobil Application

People these days have spent more time on portable devices such as their smart phone or tablet than on their computers. Researchers have recently proposed general infrastructures for portable devices within proximity of each other to trade various services in a scalable and decentralize way without going through any Internet server. Collaborating devices can synchronize their times, run localization algorithms that increase the precision of street map software, borrow each other's bandwidth, or even share cached web content. The problem with this model is that it can easily be disrupted by uncooperative and malicious sybils. Those who only want to profit from these services and not providing anything in return. They can usually create a number of sybil identities to avoid being tracked down since no identity certification authority is involved in this kind of model.[9]

IV. Sybil Defense

1) Defense Goal: Before going into the details of sybil defenses, the defense goal should be defined. It should be obvious that the ideal goal is to eliminate Sybil attack but this goal might not be realistic deal to the fact that we don't want to enforce a strict rule whenever people join the our system. Without a centralized trusted certification based scheme, the best we can do is to restrict the effects of sybils. False positive rate and false negative rate can serve as good metrics when evaluating a defense approach. False positive happens when a honest node is identified as sybil using the defense mechanism and false negative happens when a sybil node is not detected and treated as a honest user. So the goal in the following defense mechanisms is to minimize the false positive and false negative rates as much as possible.

2) Classification of Defenses: There are general defense mechanisms that work in most domains and there are domain specific approaches that target a specific domain and works better in that domain by levering some domain specific features. We further divide the general approach into traditional approach and social network based approach. Traditional approach were approaches found between 1999 to 2004. They focus mainly on building identity authentication, resource testing or human assisted sybil detection mechanisms into existing systems. On the other hand, social network based approach started around 2006 tries to detect sybils by exploiting key social network structures underlying the system.

V. General Approach

There are two types of general approaches. The traditional approach and the social network based approach.

A. Traditional Approach

Traditional approach is based on research starting around 1999 and end around 2004. During that period of time, researchers have focused on preventing Sybil attacks by involving secure mechanisms such as digital signatures and identity authentication. Other methods have also been found to increase the resource cost of a Sybil attacks such as resource testing and recurring cost. Moreover, the human assisted sybil detection approach has been widely deployed during that time.

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. [10][11][12].

2) Resource Testing and Recurring Cost: 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 client 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, MAC address. This approach in theory should work for systems that are in very low risk. Its easy to implement resource testing or recurring cost approaches but people these days can acquire a large amount of resources in a short period of time with the help of public cloud. Resource testing is mostly obsolete when an attacker can spin of hundreds of EC2 instances in a short period of time and terminate them after a few hours of attack. As of this writing, the EC2 t2.nano instance will only cost $0.0043 per hour with upfront payment.

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[13] periodically. The biggest disadvantage of computational puzzle is that it will prevent honest users with old computers from joining the system. Turing tests like CAPTCHA are also suggested as a recurring cost solution[14]. Using this approach, the cost of Sybil attacks have become more expensive but would the benefit still outweighs the cost? This approach is not recommended in high risk system for the following reasons. Computational puzzle can hold back entry level attacks but not advance attacks that leverage the public clouds computing power. How about turing tests? Is it not a lot of Online Social Network sites still use this approach? With crowdsourcing services like Amazon Mechanical Turk, turing tests can be crowdsourced at a reasonably low price.

3) Human Assisted Approach: We believe the human assisted approach has been the oldest sybil defense approach ever. This approach has been used to fight against identity theft. Even today, there is still no automatic way to identify fake identity. Online social network sites have long been using this approach to find fake accounts due the the failure in automated fake account detection. This approach starts with the honest users report a potential fake accounts. Afterwards, an account police start the manual inspection involving matching profile photos to the age or address, understanding natural language in posts, examining the friends of the user, etc. This approach is time consuming. Tuenti a Spanish social network receives on average 12,000 reports regarding fake accounts per day and only about 5% of them are indeed fake. This approach can effectively offload some of the sybil detection work to its honest users but should be kept to its minimum by combining it with another automated sybil detection technique.

B. 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.

VI. 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] L. Wang and J. Kangasharju, "Real-world sybil attacks in bittorrent mainline dht," in *Global Communications Conference (GLOBECOM), 2012 IEEE*, pp. 826–832, Dec 2012.

[3] P. Maymounkov and D. Mazières, "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, (London, UK, UK), pp. 53–65, Springer-Verlag, 2002.

[4] 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.

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

[6] 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.

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

[8] 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.

[9] D. Quercia and S. Hailes, "Sybil attacks against mobile users: Friends and foes to the rescue.," in *INFOCOM*, pp. 336–340, IEEE, 2010.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[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.