

Two papers on Sybil Attack

Kristiina Karvonen

John R. Douceur: The Sybil Attack

- Seminal paper on defining and naming a Sybil attack
 - Originally presented 2002 at *IPTPS*
 - About the author:
 - Works at Microsoft Research as a researcher in the Systems and Networking Research Group
 - Seems to like cats 😊
- (<http://research.microsoft.com/~johndo/personal.aspx>)

Paper's contribution

- Defining and naming a Sybil attack:
 - the forging of multiple identities
 - In a sybil attack, a malicious user obtains multiple fake identities and pretends to be multiple, distinct nodes in the system.
 - By controlling a large fraction of the nodes in the system, the malicious user is able to “out vote” the honest users in collaborative tasks.
 - The system must ensure that distinct identities refer to distinct entities.
 - named after the subject of a book Sybil, a case study of a woman with multiple personality disorder.

The source of the problem

- Large-scale peer-to-peer systems face security threats from faulty or hostile remote computing elements.
- To resist these threats, many such systems employ redundancy.
- However, if a single faulty entity can present multiple identities, it can control a substantial fraction of the system, thereby undermining this redundancy.
- In peer-to-peer systems it is difficult to avoid an adversary that masquerades under multiple identities, and thus appears to be many different people.
- This difficulty is at the core of performing identity based filtering in open publishing systems.

Douceur's approach

- Douceur's approach to preventing these Sybil attacks is to have a trusted agency certify identities.
- Douceur's solution is a requirement that all nodes get their nodeIDs from a central server which is responsible at least for making sure that the distribution of nodeIDs is even.
- This paper shows that, without a logically centralized authority, Sybil attacks are always possible except under extreme and unrealistic assumptions of resource parity and coordination among entities.
- In the absence of an identification authority, a local entity's ability to discriminate among distinct remote entities depends on the assumption that an attacker's resources are limited.

The impossible conditions

- This approach entails the following conditions:
 - All entities operate under nearly identical resource constraints.
 - All presented identities are validated simultaneously by all entities, coordinated across the system.
 - When accepting identities that are not directly validated, the required number of vouchers exceeds the number of systemwide failures.
- Douceur concludes by claiming that in a large-scale distributed system, these conditions are neither justifiable as assumptions nor practically realizable as system requirements, so Sybil attack will remain a possibility with such systems.

SybilGuard: Defending Against Sybil Attacks Via Social Networks

- Authors Haifeng Yu, Michael Kaminsky, Phillip B. Gibbons, Abraham Flaxman

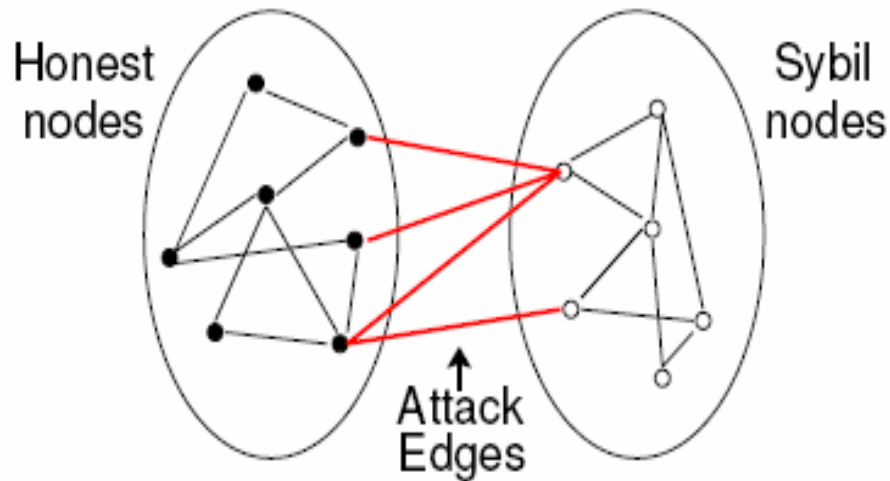
Other ways to defend against Sybil Attacks

- After Douceur's paper, p2p and other distributed systems with no central authority have been known to be vulnerable to sybil attacks
- Different approaches have been used to combat the sybil attack, two of which are subnet black listing and CAPTCHAS.
 - In the case of subnet blacklisting it is assumed that the adversary can modulate their IP address but only within a particular subsection of the IP space.
 - This belief cannot apply to a determined adversary that can (for a small fee) buy time on one of the many available botnets, spanning most of the IP space.
 - CAPTCHAS, on the other hand, are deformed strings of characters which are difficult to parse automatically.
 - They are presented to the user to make sure upon registration that a real human is indeed performing the operation, and not an automaton.
 - A typical attack against them is to relay the challenges to other users, or to simply pay others to solve them (i.e. relay them somewhere where labour is cheap).
- Neither of these approaches are suitable to defend against abuse from a determined adversary.

Paper's contribution

- This paper presents SybilGuard, a novel protocol for limiting the corruptive influences of sybil attacks.
- The protocol is based on the “social network” among user identities, where an edge between two identities indicates a human-established trust relationship.
- The main idea: Malicious users can create many identities but few trust relationships.
- Thus, there is a disproportionately-small “cut” in the graph between the sybil nodes and the honest nodes.
- SybilGuard exploits this property to bound the number of identities a malicious user can create.
- The authors try to show the effectiveness of SybilGuard both analytically and experimentally.

Basis for the design in SybilGuard



- Based on social networks
- identities are nodes in the graph and (undirected) edges are human-established trust relations (e.g., friend relations).
- The edges connecting the honest region (i.e., the region containing all the honest nodes) and the sybil region (i.e., the region containing all the sybil identities created by malicious users) are called *attack edges*.

How does it work

- if malicious users create too many sybil identities, the graph becomes “strange” in the sense that it has a small *quotient cut*—i.e., a small set of edges (the attack edges) whose removal disconnects a large number of nodes (all the sybil identities) from the rest of the graph.
- Social network do not have such cuts in practice

In practice..

- Directly searching for the cuts is not practical
- Instead, SG relies on a "special kind of verifiable random walk in the graph and intersections between such walks"
- A completely decentralized solution that with high probability guarantees that an honest node accepts connections to honest nodes only
- the number of edges of the actual social network remain the same

Basic principles of SybilGuard

- relies on properties of the users' underlying social network:
 - the honest region of the network is fast mixing
 - malicious
 - users may create many nodes but relatively few attack edges, where edges represent trust relationships