Sybil defenses via social networks

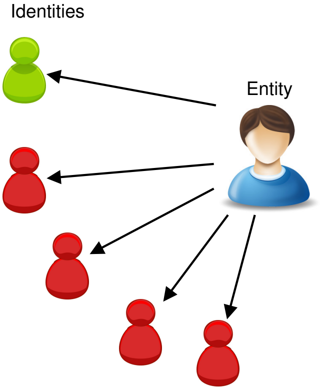
Abhishek

University of Oslo, Norway

24/04/2013

Sybil identities

- A user can pretend many fake/sybil identities
  - i.e., create multiple accounts
  - observed in real-world P2P systems
  - also observed in open systems such as Amazon
- No one-to-one correspondence between entity and identity
- Sybil identities can become a large fraction of all identities



Sybil attack

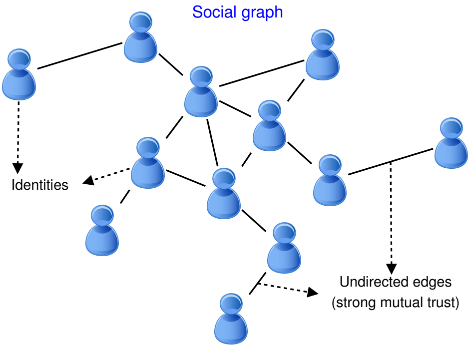
Enables malicious users to out-vote honest users

- Majority voting: cast more than one vote
- Byzantine consensus: exceed the 1/3 threshold
- DHT: control large portion of the ring
- Recommendation systems: manipulate the recommendations

Defending against sybil attacks

- Requires binding an entity to an identity
  - Difficult in absence of trusted central authority [Douceur 2002]
- Simple sybil defenses include
  - CAPTCHAs
  - IP address filtering
  - Computational puzzles
- Simple defenses
  - leave out large number of honest users, or
  - are too weak to deter resourceful attacker

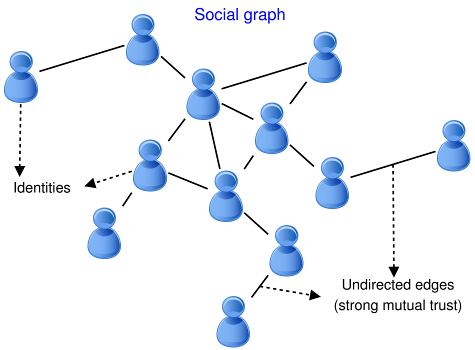
Social graph



System model

- Social graph $\mathcal{G}$
- n honest users with single honest identity
- Multiple malicious users, each with multiple identities (sybil nodes)
- Assumption: Neighbors in social graph share secret symmetric keys

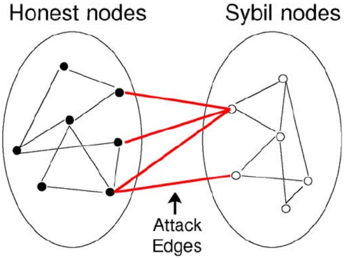
Goal of sybil defense



Goal

- Allow any given honest identity V to label any other given identity S as either honest or sybil
- Bound the total number of false negatives below the tolerance threshold of the distributed system
- Small fraction of false positives can be tolerated

Insights for SybilLimit solution



Key insights

- Assumption: The number of attack edges is independent of the number of sybil identities
- Assumption: The cut along the attack edges will have a small quotient
  - i.e., $\frac{number\ of\ attack\ edges}{number\ of\ nodes\ disconnected}$ is small
- Break symmetry to properly label nodes

General approach for the SybilLimit solution

Given an honest node V, search for a subgraph $\mathcal{H}$ of $\mathcal{G}$ such that

- $\mathcal{H}$ contains V
- $\mathcal{H}$ has n nodes (n: number of honest nodes in system)
- the minimum quotient cut of $\mathcal{H}$ is not excessively small

Challenge

Make sure that $\mathcal{H}$ does not grow in sybil region

From cuts to mixing time

Difficulty with cuts

- Need to perform computation over all nodes
- Centralized

Idea

If a subgraph has small quotient cut, then the mixing time of the subgraph is large

Advantage

- Can be performed in incremental manner
- Decentralized

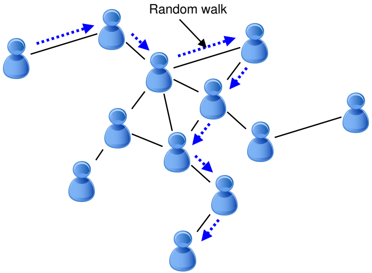
Mixing time

- Stationary distribution of a random walk
  - a probability distribution π that is invariant to the transition matrix P
  - i.e., $\pi P = \pi$
- Mixing time of a random walk, T
  - minimal length of the random walk in order to reach the stationary distribution

$$P = \begin{pmatrix} 0 & \frac{1}{deg(v_1)} & \cdots & 0 \\ \frac{1}{deg(v_2)} & 0 & \cdots & \frac{1}{deg(v_2)} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{deg(v_n)} & 0 & \cdots & 0 \end{pmatrix}$$

$$\pi = \begin{bmatrix} \frac{deg(v_1)}{2m} & \frac{deg(v_2)}{2m} & \cdots & \frac{deg(v_k)}{2m} \end{bmatrix}$$

m : no. of edges in the undirected graph



Assumption for mixing time of $\mathcal{H}$

Assumption for SybilLimit

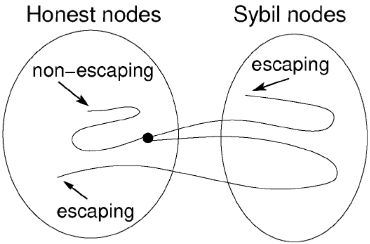
The honest region (i.e. subgraph) of $\mathcal{G}$ has a mixing time no larger than $t(n)$, where $t(n)$ is a function of the size n of the honest region

- SybilLimit assumes $t = \mathcal{O}(\log n)$
- Theoretical evidences exist to support $t = \mathcal{O}(\log n)$ for some models of social networks such as Kleinberg's social network model

Solution basis

Use random walks in $\mathcal{G}$ to exploit its abnormal mixing time for differentiating sybil nodes from honest nodes

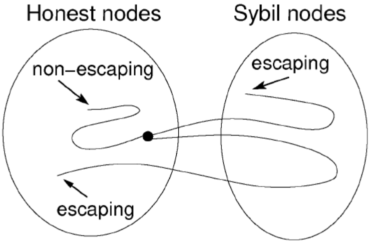
Escaping random walks and escaping nodes



Probability for escaping random walks

- Escaping probability of a length-w random walk starting from a uniformly random honest node is at most gw/n
  - g: total number of attack edges
  - n: total number of honest nodes
  - assumes honest nodes form a connected component

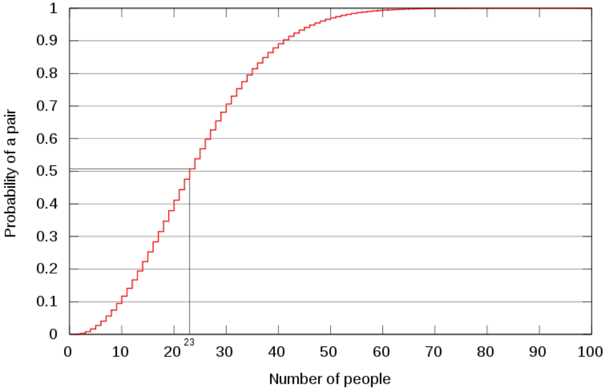
Escaping random walks and escaping nodes



Only protects non-escaping nodes

- For at most ε fraction of honest nodes, corresponding probability is above $(gw)/(n\varepsilon)$
- Provable guarantees only for non-escaping nodes
- Escaping nodes are likely to be close to attack edges

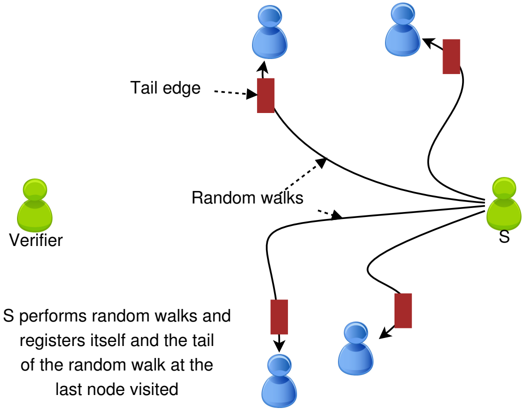
Birthday paradox



Approximate probability of at least two people sharing a birthday amongst a certain number of people

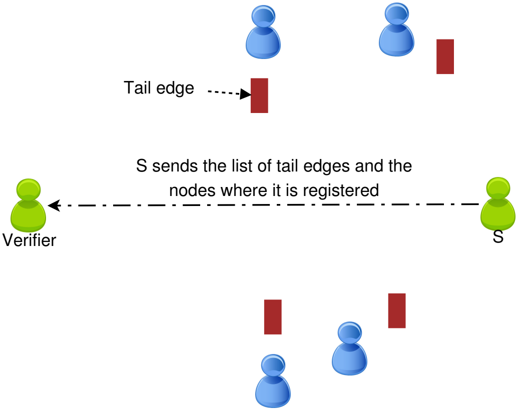
- Assumes all birthdays are equally likely
- In the honest region, all edges are equally likely to be tail of random walks

SybilLimit protocol in honest region



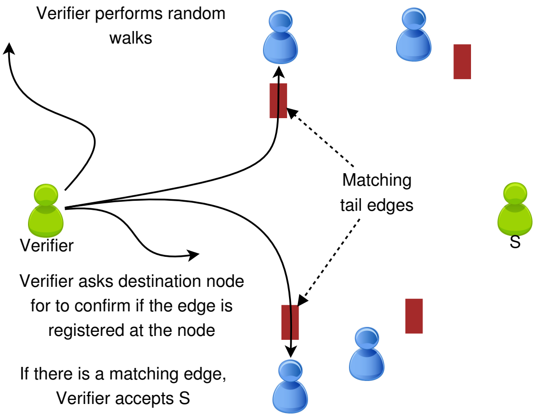
- Scenario when both Verifier and S are in honest region
- S performs $\Theta(\sqrt{m})$ random walks where m is the number of edges in honest region
- $\Theta(\sqrt{m})$ random walks results in $\Theta(\sqrt{m})$ tail edges

SybilLimit protocol in honest region



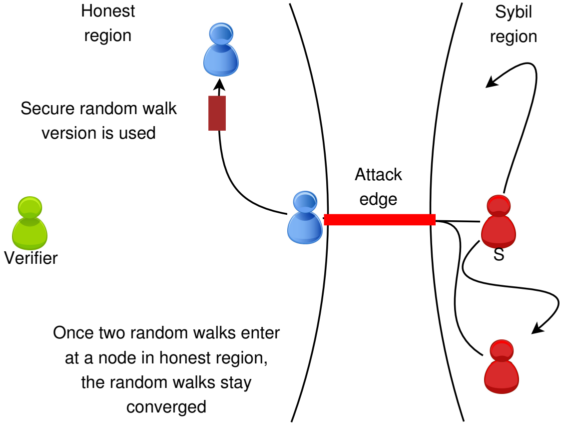
- Scenario when both Verifier and S are in honest region
- S performs $\Theta(\sqrt{m})$ random walks where m is the number of edges in honest region
- $\Theta(\sqrt{m})$ random walks results in $\Theta(\sqrt{m})$ tail edges

SybilLimit protocol in honest region



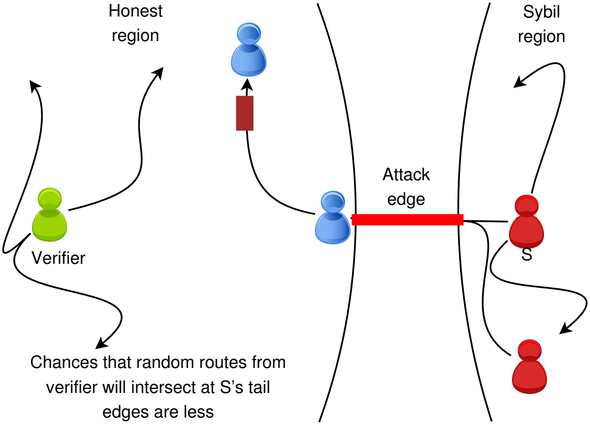
- All edges in honest region are equally likely
- By Birthday paradox, there is a high probability that a matching edge is found by the Verifier

SybilLimit protocol when sybil nodes are involved



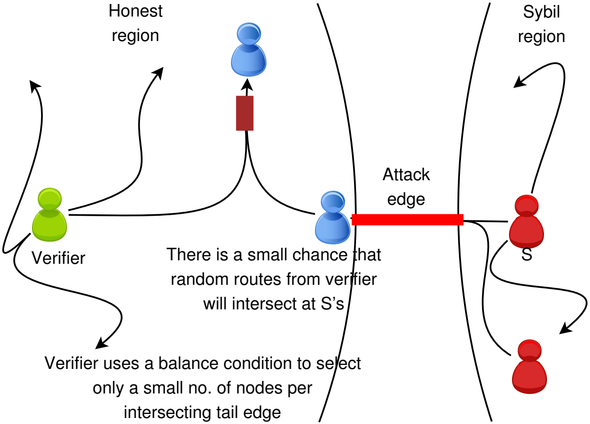
- Scenario S is in sybil region
- A node uses a tail to label only $\Theta(n/\sqrt{m})$ nodes
- For sybil nodes collectively, the umber of possible tainted tails is bound within $O(gt\sqrt{m})$

SybilLimit protocol when sybil nodes are involved



- Scenario S is in sybil region
- A node uses a tail to label only $\Theta(n/\sqrt{m})$ nodes
- For sybil nodes collectively, the umber of possible tainted tails is bound within $O(gt\sqrt{m})$

SybilLimit protocol when sybil nodes are involved



- Scenario S is in sybil region
- A node uses a tail to label only $\Theta(n/\sqrt{m})$ nodes
- For sybil nodes collectively, the umber of possible tainted tails is bound within $O(gt\sqrt{m})$

SybilLimit results

Formal guarantees

- Assuming
  - Honest region has mixing time no larger than t
  - Number of attack edges, $g = o(n/t)$
- A honest node V with probability at least $1-\delta$ (for $\delta > 0$) labels
  - At least $(1-\varepsilon)n$ honest nodes as honest (for $\varepsilon > 0$)
  - At most $O(t)$ sybil nodes per attack edge as honest

Numerical example

- Sample social network sizes: 100,000 to 1,000,000 honest nodes
- Sybil nodes generated synthetically
- Labels 95% of honest nodes as honest
- Labels 10-20 sybil nodes as honest per attack edge

Estimating unknown parameters

t: size of random walks (mixing time)

- In practice, $t = O(\log n)$
- Simply use a t around 20 or 30 (sufficient for 1 million nodes)
- Increasing t linearly increases the number of false negatives

m: total number of edges in honest region

- Estimates m using a benchmarking technique
- Never over-estimates, but under-estimation is possible

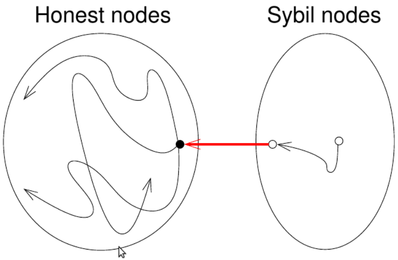
Practical implication and deployment considerations

- What can be used as a social network for sybil defense?
- Do social networks have really small mixing time?
  - Research community divided into 2 camps
  - SybilLimit removes low-degree nodes while performing evaluation
  - Do we really need small mixing time?
- Will targeted sybil attacks break these defenses in practice?

SybilLimit summary

- Sybil defense mechanisms via social networks
- Assumptions
  - Social graph has low mixing time for random walks
  - Sybil nodes can establish only a small number of edges with honest nodes
- Exploits the knowledge that addition of sybil nodes increases the mixing time for random walks
- Allows a node to identify another node as honest or sybil
- Decentralized

Secure random walks



- Sybil nodes can perform unlimited number of random walks
- These may result in large number of different tainted tails
- Possible solution
  - Each edge in the graph enforces a quota on the total number of times that edge can be crossed by all random walks collectively

Secure random walks using random routes

- Random routes in place of random walks
  - Rather than selecting next hop randomly, there is a random mapping between incoming edge and outgoing edge
- Each node maintains $\Theta(m)$ independent instances of routing table
- If two random routes in a given instance ever cross the same edge, they merge and stay together for ever
- If a random route encounters some node more than once, that node will use additional independent routing tables for those extra routing decisions
- Node keeps track of hop count viewed for a random routes
- Node can drop random routes when they observe that hop counts are not maintained correctly

Comparison of social-based social defenses

Protocol	*Assumption*	*Main technique*	*Provable end guarantee?*	*Complete decentralized design?*
SybilGuard [30] and SybilLimit [27, 28]	Assumption 1	random walk	√	√
SybilInfer [8]	Assumption 1	random walk	×	×
Gatekeeper [23]	Assumption 1 and 2	breadth-first search and random walk	√	√
SumUp [24]	Assumption 1*	adaptive max flow	√	×
Applying community detection algorithms [25]	not clearly made, but likely similar to Assumption 1	detecting social communities	×	×
Whanau [16]	Assumption 1	random walk and layered-IDs for DHT	√	√
Ostra [18]	not clearly made, but likely similar to Assumption 1	"remove" certain edges based on user feedback	×	×

*Note that [24] only mentions Assumption 1, but we are not sure whether a similar assumption as in Gatekeeper is needed.

- **Assumption 1:** The honest region (i.e., subgraph) of $\mathcal{G}$ has mixing time no larger than $O(\log n)$
- **Assumption 2:** The honest region (i.e., subgraph) of $\mathcal{G}$ is reasonably balanced

Comparison of defenses against sybil attacks not based on social graph

Techniques	Advantages	Disadvantages
Certificates signed by trusted authority (Castro et al. 2002)	- Controls who can join the system	- Administrative overhead - Certificate revocation may be costly
Distributed registration (Dinger et al. 2006)	- No barriers to enter - Decentralized	- Fails under attacks involving large no. of IPs - New attacks possible
Use of bootstrap graph based on social network (Danezis et al. 2005)	- No barriers to enter - Decentralized	- Significant overhead - Not sure if it scales

Comparison of defenses against sybil attacks not based on social graph. . .

Techniques	Advantages	Disadvantages
Use of physical network characteristics to identify nodes (Wang et al. 2005)	- No barrier to enter	- Lack of consistent identity resulting from change in measurement over time - Changes to the network measurement infrastructure may invalidate the identity of all nodes
Use of network coordinates to group nodes (Bazzi et al. 2005)	- Works when a single node is reporting multiple identities - Works when a group of nearby nodes are colluding	- Fails when attacker controls large number of nodes in multiple network positions - May require a trusted network measurement infrastructure
Use of network coordinates to differentiate nodes (Bazzi et al. 2006)	- Hop-count distance used to tell physically nodes separated nodes apart	- Fails when attacker controls large number of nodes in multiple network positions - Requires appropriately placed trusted beacons

Comparison of defenses against sybil attacks not based on social graph…

Techniques	Advantages	Disadvantages
Computational puzzles (Borisov 2006)	- Works for computationally limited adversaries - Decentralized	- Overhead for honest nodes - Difficult to choose appropriate puzzle - Nodes can choose their ID, which facilitates targeted attacks
Computational puzzles generated hierarchically (Rowaihy et al. 2007)	- Works for computationally limited adversaries	- Requires centralized online trusted authority - Requires reliable nodes in the upper levels of the certification hierarchy
Economic incentives (Margolin et al. 2007)	- Decentralized	- Requires implementation of currency - Requires expressing all costs and utilities in terms of a currency - Only detection of attack

References

1. John R. Douceur. 2002. The Sybil Attack. In Revised Papers from the First International Workshop on Peer-to-Peer Systems (IPTPS '01)
2. Abedelaziz Mohaisen, Aaram Yun, and Yongdae Kim. 2010. Measuring the mixing time of social graphs. In Proceedings of the 10th annual conference on Internet measurement (IMC '10). ACM, New York, NY, USA, 383-389.
3. Haifeng Yu. 2011. Sybil defenses via social networks: a tutorial and survey. SIGACT News 42, 3 (October 2011), 80-101.
4. Haifeng Yu, Phillip B. Gibbons, Michael Kaminsky, and Feng Xiao. 2010. SybilLimit: a near-optimal social network defense against sybil attacks. IEEE/ACM Trans. Netw. 18, 3 (June 2010), 885-898.
5. Guido Urdaneta, Guillaume Pierre, and Maarten Van Steen. 2011. A survey of DHT security techniques. ACM Comput. Surv. 43, 2, Article 8 (February 2011), 49 pages.