SybilInfer: Detecting Sybil Nodes using Social Networks

Ashutosh Jain

(2011MCS2566)

Chandra Prakash

(2011MCS2610)

Motivation

- A single entity/user can pretend to have multiple identities
    - Sybil Attack
- Distributed Systems Security
    - Byzantine Consensus
    - Secure routing in DHTs
- SybilInfer is an algorithm for labelling nodes in a social network as honest user or Sybils.
- Assumption : bound on fraction of dishonest identities

Sybil Attack

- Sybil identities can own a large fraction of all identities

  -Distributed systems security solutions fail...

- Botnets

  -Zombie machines

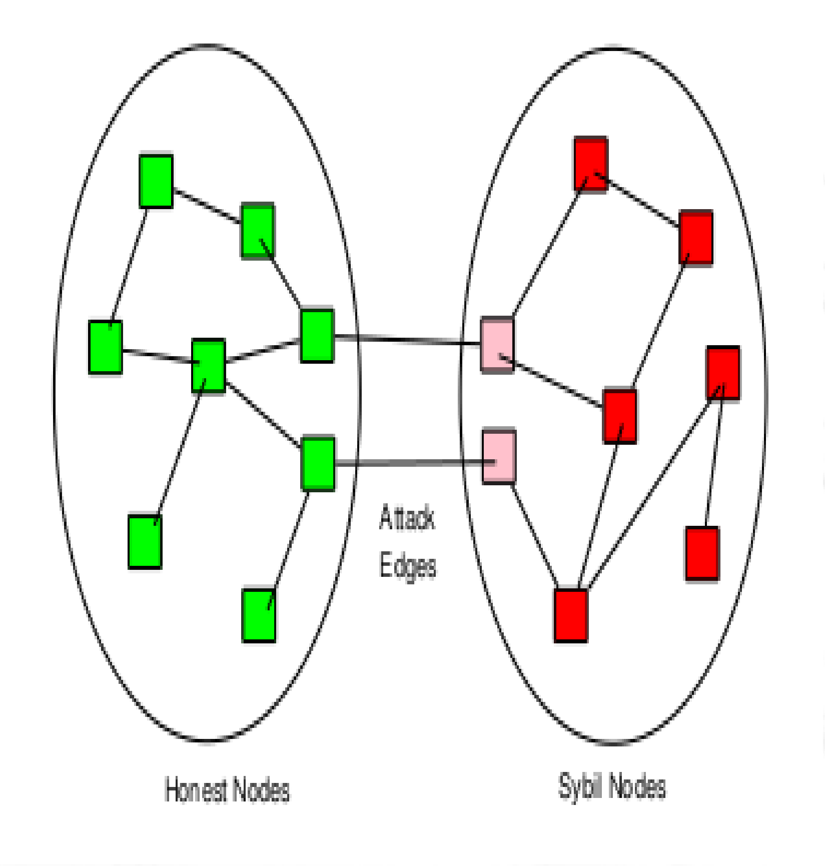
  -Average size > 20,000

How to bound the fraction of dishonest nodes?

- Trusted Central authority
- Distributed Solutions?
- Social Networks

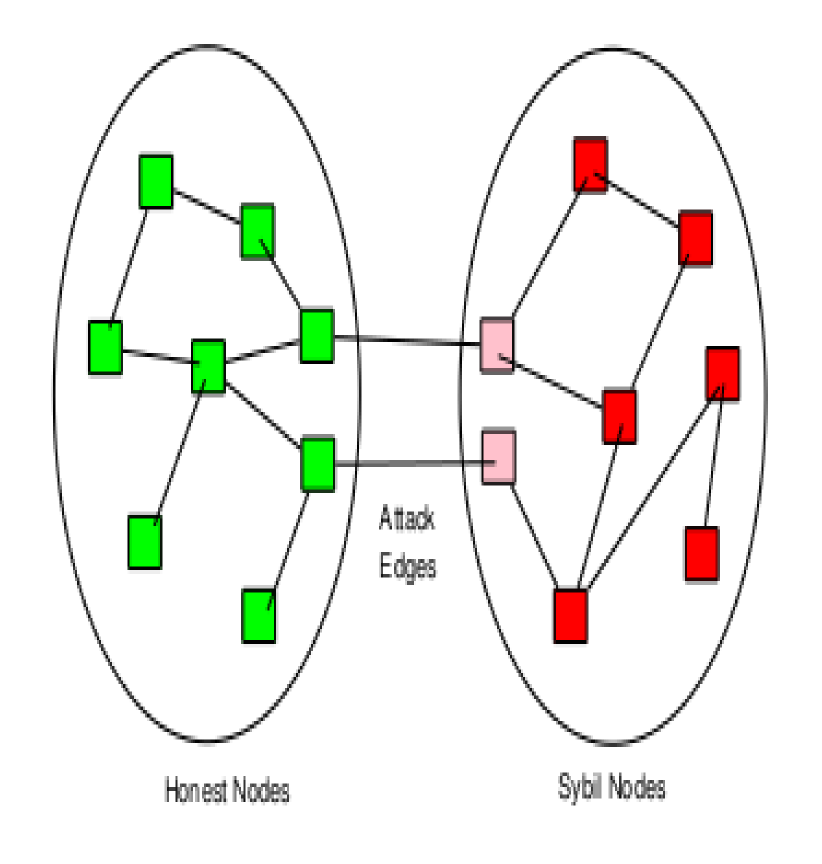
Leveraging Social Networks

- Resource Constraint
  - bound on number of trust relationships between attackers and honest nodes
  - Attacker cannot create edges between honest nodes and Sybil identities



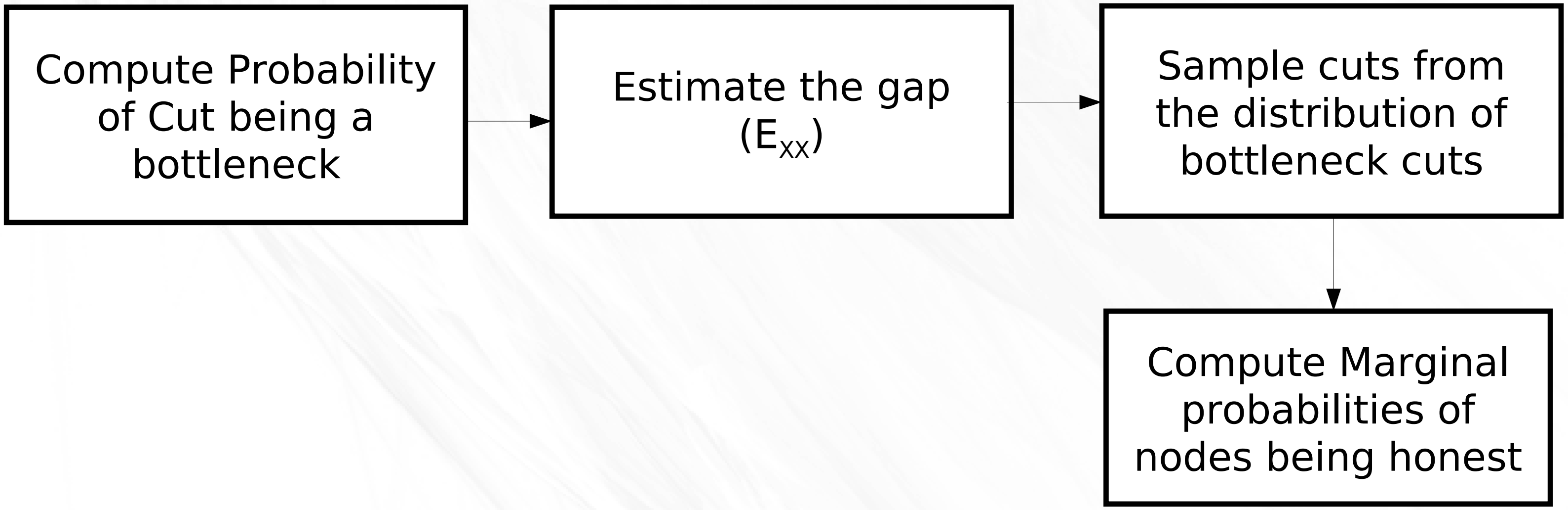
Leveraging Social Networks

- Social networks are Fast Mixing
  - Random walks quickly convergence to stationary distribution
- Sybil attacks induce a bottleneck cut
  - Fast mixing is disrupted
- Knowledge of an apriori honest node
  - Breaks Symmetry



Approach used:

- Design Philosophy
  - Optimal use of all information available in the graph
  - No assumptions on threshold of attack edges



Compute Probability of Cut being a bottleneck

Estimate the gap (E_{XX})

Sample cuts from the distribution of bottleneck cuts

Compute Marginal probabilities of nodes being honest

Formal Model

- Properties of Mixing times
  - Depend on random walks
  - and where they end
- Each vertex performs S random walks
  - length l =log(|V|)
  - Transition probability

$$P_{ij} = \begin{cases} \min\{\frac{1}{d_i}, \frac{1}{d_j}\} & \text{if i} \longrightarrow \text{j} \in \text{E} \\ 0 & \text{otherwise} \end{cases}$$

  - Uniform stationary distribution (without attack)
- Let T = set of vertex pairs <start vertex, end vertex> for each random walk called Traces.

Formal Model

- Assign probabilities of cuts being honest

$$P(X = Honest \mid T)$$

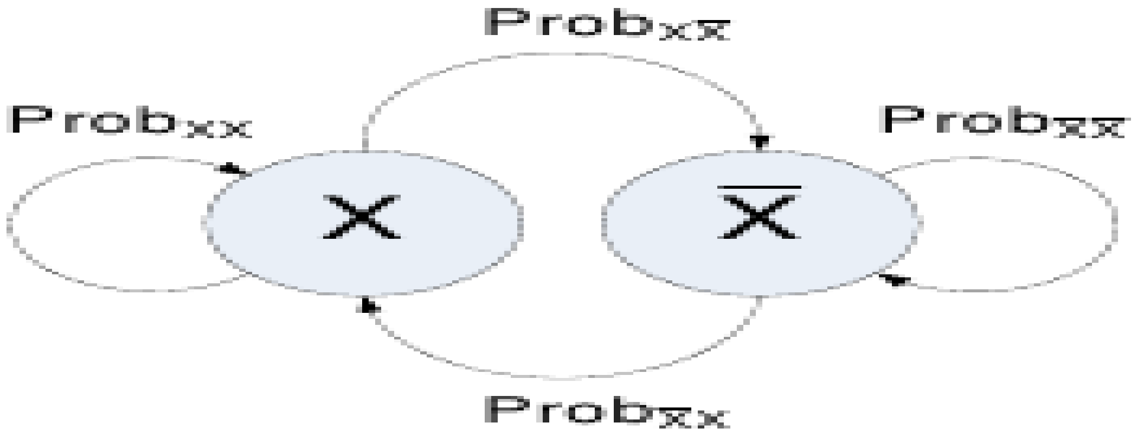
- Using Bayes Theorem, we have that :

$$P(X = Honest \mid T) = \frac{P(T \mid X = Honest) \cdot P(X = Honest)}{Z}$$

$$Z = \sum_{X \subset V} P(T \mid X = Honest) \cdot P(X = Honest)$$

- Next Challenge: Model $P(T \mid X = Honest)$

Formal Model


$\mathrm{Prob}_{X\overline{X}}$
Prob_{XX}
$\mathrm{Prob}_{\overline{X}\overline{X}}$
X
$\overline{X}$
$\mathrm{Prob}_{\overline{X}X}$

$$prob_{XX} = \frac{1}{|V|} + E_{XX}$$

$$prob_{X\overline{X}} = \frac{1}{|V|} - E_{X\overline{X}}$$

$$P(T \mid X = honest) = \left(prob_{XX} \right)^{N_{XX}} \left(prob_{X\overline{X}} \right)^{N_{X\overline{X}}} \left(prob_{\overline{X}\overline{X}} \right)^{N_{\overline{X}\overline{X}}} \left(prob_{\overline{X}X} \right)^{N_{\overline{X}X}}$$

Estimating E_{XX} / $prob_{xx}$

- We could sample E_{XX} as well
  - $P(X,E_{XX}|T)$
  - Expensive
- Instead, we shall directly estimate the best E_{XX}

$$prob_{xx} = \frac{\sum_{x\in X}\sum_{y\in X} P_{xy}{}^{l}}{|X|} \cdot \frac{1}{|X|}$$

$$prob_{xx} = \frac{N_{XX}}{N_{XX} + N_{X\overline{X}}} \cdot \frac{1}{|X|}$$

$$P(T \mid X = Honest)$$

$$P(T \mid X = H) = \left(\frac{N_{XX}}{N_{XX} + N_{X\overline{X}}} \cdot \frac{1}{|X|} \right)^{N_{XX}} \left(\frac{N_{X\overline{X}}}{N_{X\overline{X}} + N_{XX}} \cdot \frac{1}{|\overline{X}|} \right)^{N_{X\overline{X}}} \left(\frac{N_{\overline{X}\overline{X}}}{N_{\overline{X}\overline{X}} + N_{\overline{X}X}} \cdot \frac{1}{|\overline{X}|} \right)^{N_{\overline{X}\overline{X}}} \left(\frac{N_{\overline{X}X}}{N_{\overline{X}X} + N_{\overline{X}\overline{X}}} \cdot \frac{1}{|X|} \right)^{N_{\overline{X}X}}$$

Sampling

$$P(X = Honest \mid T) = \frac{P(T \mid X = Honest) \cdot P(X = Honest)}{Z}$$

- Sample from above distribution
- Marginal Probabilities
  - P(Node j is honest) = # j appears in samples/ #samples
  - Can label nodes as honest/dishonest
- Sampling algorithm : Metropolis-Hastings
  - Current State : X_0
  - Propose a new state X_1 with probability $Q(X_1|X_0)$
  - Accept new state with probability

$$\min\{\frac{P(X_1 = Honest \mid T) Q(X_0 \mid X_1)}{P(X_0 = Honest \mid T) Q(X_1 \mid X_0)}, 1\}$$

Theoretical Guarantees

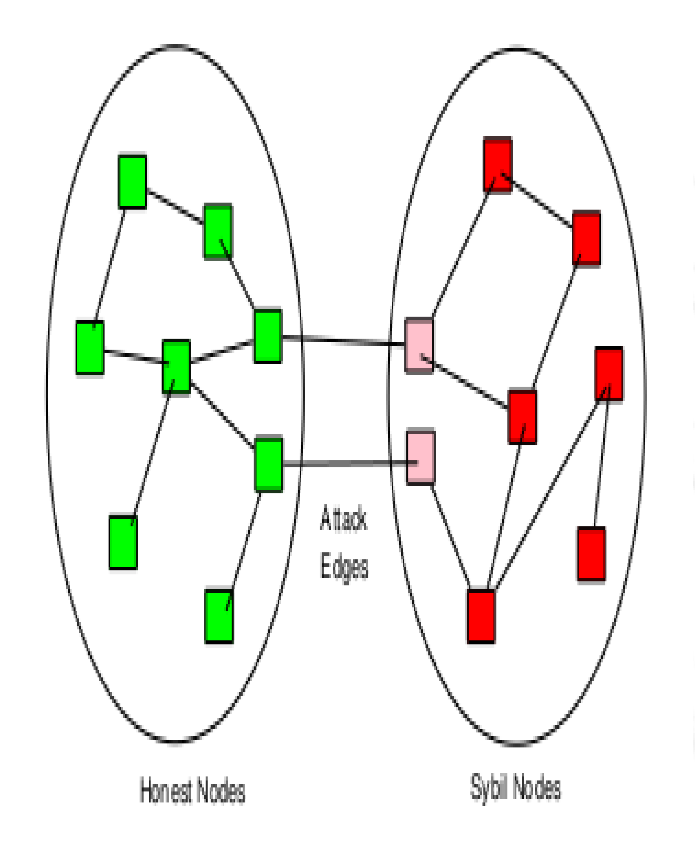
- Ideal Scenario:
  - Without attack, the cuts obtained from model have E_{XX}=0
  - Under attack, the cuts obtained from the model have E_{XX} >0 regardless of attacker strategy
- Real World:
  - Without attack, we obtain cuts with E_{XX} approx 0 (upper bounded by E_{max})
  - Under a major Sybil attack, we obtain cuts with $E_{XX} > E_{max}$ regardless of attacker strategy

LiveJournal

- Extract a social sub graph from LiveJournal
  - Three hop neighbourhood of a random node
- Processing
  - Remove nodes with degree < 3
  - 33170 nodes
- The model found a bottleneck cut is this topology
  - False positive or Sybil attack?
  - Remove the bottleneck cut
  - 31603 nodes

Related Work

- SybilGuard[SIGCOMM 06] & SybilLimit [Oakland 08]
  - Assumes short random walks lie mostly in the honest region
    - Results in poor threshold to colluding attackers
  - Heuristic validation approach
    - Honest nodes random walks intersect
    - Birthday paradox
    - High false negatives



Conclusions

- Proposed a formal model for inferring Sybil identities in a Social Network
- Proposed solution can be applied to security critical centralized/distributed applications
  - High tolerance to colluding adversary
  - Low false negatives

References

1. G. Danezis and P. Mittal. Sybilinfer: Detecting Sybil nodes using social networks. In NDSS, 2009.

Thank you!

Questions??