Sybil-resilient online content voting

Nguyen Tran, Bonan Min, Jinyang Li, and Lakshminarayanan Subramanian

New York University
NSDI’09
Sybil attacks pollute voting result

Websites rely on votes for ranking

The Best Dam Water Money can Buy

failedsuccess.com — A new product is sweeping the nation, breathing life into the industry.

REALLY Cheap Web Host

mher.org — Plans starting at $0.99; 8000 MBs of space and 150 GBs of bandwidth. Very fast and reliable. They never oversell. You can see examples of how feblesiaic.com and slackpacks.org.

Free Ipod Shuffle?

FreePodShuffle.com — MAC has just introduced a new Ipod to the bunch. Ipod Shuffle is going to be a HOT item this year!

Web site attacks Sybil attacks pollute voting result. Websites rely on votes for ranking.
Sybil defense is hard

• Open system → an attacker can join easily
• Lack of strong identity → an attacker can join with many fake accounts

• Need some resource that cannot be acquired in abundance
  – Links in a social network?
Social links are hard to acquire in abundance

Assumption: # of attack edges is small

[SybilLimit S&P’08] [Ostra NSDI’08]

An attack edge connects an honest node and a Sybil node

[SybilGuard Sigcomm’06]
SumUp: a Sybil-resilient vote aggregation system

- SumUp’s setting: A central party collects all votes and the social graph
- Goal: extract a subset of votes
  - Include few votes from Sybils
  - Include most votes from honest users
SumUp: a Sybil-resilient vote aggregation system

- SumUp’s setting: A central party collects all votes and the social graph
- Goal: extract a subset of votes
  - Include few votes from Sybils
  - Include most votes from honest users
SumUp overview

• Design insights
  1. Designate a vote collector
  2. Use max-flow to collect votes
  3. Assign appropriate link capacities
#1: Designate a vote collector

- Each node acts as its own vote collector
#2: Use max-flow to collect votes

Bogus votes are congested at attack edges
#2: Use max-flow to collect votes

vote collector

Honest votes are also congested

Bogus votes are congested at attack edges
#3: Assign appropriate link capacities

- Attack edges should be farther away from the collector.
- Honest votes are congested at edges closer to the collector.
SumUp’s design details

- Capacity assignment
- Leverage user feedback to reduce bogus votes
Capacity assignment

- Assign link capacity to collect at most $v$ votes
  - Ideally $v$ is the number of honest votes

- Assign greater capacity to edges that are closer to the vote collector
Assign capacity via ticket distribution

A node consumes 1 ticket and distributes the rest to out-going links

$v = 6$ tickets
Assign capacity via ticket distribution

Link capacity = (# of tickets on the link) + 1

v = 6 tickets
Assign capacity via ticket distribution

Link capacity = (# of tickets on the link) + 1

v = 6 tickets
Approximate $v$ to # of honest votes

• Observation
  – When number of honest votes $\gg v$, number of collected votes $\approx v$
  – When number of honest votes $\ll v$, number of collected votes $\ll v$
  • Not many bogus votes are collected
Setting $v$

- Iteratively adjust $v$:

1. Start with a small $v = 100$
2. Collect votes using current $v$
3. If # of collected votes > $0.5^*v$, double $v$ and repeat step 2

- Final $v$ approximates number of honest votes regardless of attacks
SumUp’s provable properties

• Limit bogus votes
  – When $v \ll n$, expected bogus votes per attack edge is $1+ o(1)$
  – Even when $v = \Theta(n)$, expected bogus votes per attack edge is $O(\log n)$ [SybilLimit IEEE S&P’08]

• Collect a large fraction of honest votes
  – On a random graph, $\sim 100\%$ honest votes can be collected
Leverage user feedback on votes

• If vote collector can tag some votes as bogus, SumUp can do better:
  – Reduce capacity on attack edges close to the collector
  – Possibly ignore attack edges
• Idea: penalize all links along the path taken by the bogus vote [Ostra NSDI’08]
Associate a penalty with each link

An attack edge is close to the vote collector
Reassign capacity according to penalty

\[ V = 6 \text{ tickets} \]

Links with higher penalty receive fewer tickets
Capacity of attack edge is reduced from 2 to 1

Reassign capacity according to penalty

V = 6 tickets
Eliminate links with high penalty

$V = 6$ tickets
Evaluation

1. How does SumUp perform on real social networks?
2. Can SumUp detect Sybil attacks?
Simulation setup

• Use 3 social networks
  – YouTube (0.5 million nodes)
  – Flickr (1.5 million nodes)
  – Synthetic (3 million nodes)
• Inject 100 attack edges randomly
• Use a random vote collector
• Choose a random set of honest voters
SumUp limits number of bogus votes

- Avg bogus votes per attack edge
- Fraction of honest nodes that vote

When all nodes vote, bogus votes per attack edge is still small

Bogus votes per attack edge is close to 1
SumUp collects most honest votes

SumUp can collect >90% honest votes
Digg marks “popular” articles
- 130,000 popular articles among 7 million articles submitted

Users cast positive or negative votes

Digg’s social network
- 3 million nodes
- 0.5 million nodes in a connected component
- 80% of votes are from the connected component
Evaluating SumUp on Digg

• Kevin Rose (Digg founder) → vote collector
• Run SumUp for all votes cast before the article is marked as “popular”

• Normal articles → fraction of votes collected > 0.7
• Suspicious articles → fraction of votes collected ≪ 0.7
Suspicious articles have evidence of attack

• ~800 suspicious articles have less than 50% votes collected by SumUp
• Manually sampled 30 articles
• Found (subjective) evidence of attacks in 15 the articles
  – 1 article is an advertisement
  – 10 articles have many newly registered voters
  – 4 articles receive < 50 votes after marked “popular”
Examples of suspicious articles

**The Best Dam Water Money can Buy**
failedsuccess.com — A new product is sweeping the nation, breathing life into the industry.

**REALLY Cheap Web Host**
mher.org — Plans starting at $0.99; 8000 MBs of space and 150 GBs of bandwidth. Very fast and reliable. They never oversell. You can see examples of how fast they are and what they can do.

**Access Your PC Anywhere and Anytime for FREE**
webex.com — With MyWebEx PC, you can access your files, programs and email from your PC at the home or office with ease, and it’s 100% Free and No Credit Card Required.

**Free Ipod Shuffle?**
FreePodShuffle.com — MAC has just introduced a new Ipod to the bunch. How about a Free Ipod Shuffle is going to be a HOT item this year!
An example of suspicious articles with no evidence of attack
Suspicious articles receive more negative votes

- Obtained negative votes for 5794 “popular” articles from 08/2008 to 09/2008
Related work

- **Node admission**
  - SybilGuard [Sigcomm’06], SybilLimit [IEEE S&P’08], SybilInfer [NDSS’09]

- **Fighting spam**
  - Ostra [NSDI’08]

- **User reputation systems**
  - SybilProof [P2PEcon’05], Appleseed [ISF’05], Advogato [SSYM’98]

- **Content voting systems**
  - Credence [NSDI’06]
Conclusion

- Defending against Sybil attacks is important for content voting systems
- SumUp vote aggregation:
  1. Limit # of bogus votes by # of attack edges
  2. Collect many votes from honest users
  3. Ignore votes from attackers that repetitively cast bogus votes