Reputation Systems II

Sybil Attack, BlogRank, B2Rank, EigenRumor, MailRank, TrustRunk

Yury Lifshits

Caltech

http://yury.name

Caltech CMI Seminar
March 4, 2008

Outline

1. Sybil Attack
2. Ranking Blogs
3. Reputations For Fighting Spam
4. Conclusions

1

Sybil Attack

Sybil Attack

- Graph of trust-weighted edges
- n honest nodes + adversary
- overall trust value on attack edges (honest-malicious) is limited

Question: whether splitting adversarial node into many is beneficial for acquiring higher reputation (rank)?

Negative Result

Assume reputation scores remain the same under isomorphism.
Is it sybilproof?

Unfortunately, no. Attack strategy?

Answer: double the graph.

Positive Results (1/3)

General form of trust flow reputations:

$$r(x) = \max_{\mathcal{P}_{tx}} \bigoplus_{p \in \mathcal{P}_{tx}} trust(p)$$

Notation:

- t is pre-trusted node
- $\mathcal{P}_{xy}$ is a family of disjoint paths from t to x

Positive Results (2/3)

Assumptions:

1. Extending path nonincreases the $trust(p)$
2. $\bigoplus$ and $trust$ are monotone to number of paths and edges values, respectively
3. Splitting a path into two does not increase $\bigoplus$ value
4. $\bigoplus = \max$

Positive Results (3/3)

Under assumptions (1-3) sybil attack does not increase adversary's reputation

Under assumptions (1-4) sybil attack does not increase adversary's rank

Proof?

SybilGuard (1/2)

- Assume number of attack edges is $A = o(\sqrt{n}/\log n)$
- System is distributed, honest nodes follow the same protocol
- Can an honest node t identify (w.h.p.) $2A+1$ nodes in such a way that at most A of them are powered by adversary?

SybilGuard (2/2)

- For every node fix a bijective mapping from in-edges to out-edges
- Take a walk from t of length at most $\sqrt{n}\log n$ using bijection routing
- At some point make a random switch, than continue another $\sqrt{n}\log n$ steps using backwalk routing
- Report a point. Repeat, until $2A+1$ points are collected

Claim

w.h.p. at most A reported nodes are malicious

2

Ranking Blogs

Ranking Blogs: Factors

- Entities: blogs, posts, communities, comments, brand names, external websites
- Frineds, blogroll, subscriptions, hyperlinks, visitors, clicks, votes
- Time
- Tags

BlogRank

Any ideas how to rank blogs?

Why not just PageRank?

Wait a minute, for which graph? Linked blogs:

- Hyperlinks, blogrolls
- Common commentors/authors, tags, co-references to news

B2Rank

$B2Rank(x) = BlogReputation \times PostQuality$

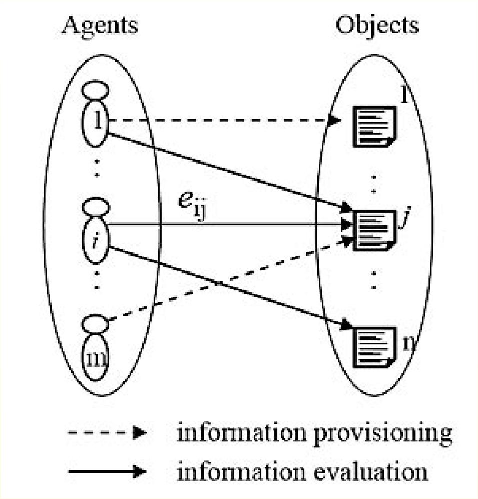
BlogReputation is computed in PageRank style for blogroll graph with one change:

- Blogroll links are weighted by activity level (frequency of blogging and commenting)

PostQuality is average for PageRank-style score of blog posts

- Post-to-post links are weighted by referring post activity and time difference

EigenRumor (1/2)



Picture from “The EigenRumor Algorithm for Ranking Blogs” paper

EigenRumor (2/2)

Notation:

- $\bar{r}$: reputation score for posts
- $\bar{a}, \bar{h}$: authority and hub scores for bloggers
- P, E: provision and evaluation matrices

$$\bar{r} = \alpha P^T \bar{a} + (1 - \alpha) E^T \bar{h}$$
$$\bar{a} = P\bar{r}, \quad \bar{h} = E\bar{r}$$

Solution: iterative algorithm for $\bar{r}$:
$\bar{r} = (\alpha P^T P + (1 - \alpha) E^T E)\bar{r}$

3

Reputations For Fighting Spam

Combining Two Scores

- Hyperlink graph
- Pre-trusted nodes
- Spam nodes
- Reputation propagates in a forward manner
- Spam score propagates backwards
- Compute spam scores a-la PageRank
- Reweight hyperlink graph and pre-trusted nodes
- Compute reputations a-la PageRank

4

Conclusions

Challenges

- Measurable objectives?
- Model for input data?
- Dynamic aspects of reputations? Digg-style ranking?
- Price of attack?
- Ranking in social networks?
- Ranking in RDF data?
- Billion dollar question: how to avoid arms race?

References

K. Fujimura, T. Inoue, M. Sugisaki
The EigenRumor Algorithm for Ranking Blogs

A. Kritikopoulos, M. Sideri, I. Varlamis
BlogRank: ranking weblogs based on connectivity and similarity features

M.A. Tayebi, S.M. Hashemi, A. Mohades
B2Rank: An Algorithm for Ranking Blogs Based on Behavioral Features

A. Cheng, E. Friedman
Sybilproof reputation mechanisms

H. Yu, M. Kaminsky, P.B. Gibbons, A, Flaxman
SybilGuard: defending against sybil attacks via social networks

P.A. Chirita, J. Diederich, W. Nejdl
MailRank: using ranking for spam detection

Z. Gyongyi, H. Garcia-Molina, J. Pedersen
Combating web spam with TrustRank

M. Dalal
Spam and popularity ratings for combating link spam

http://yury.name
http://yury.name/reputation.html
Ongoing project: http://businessconsumer.net

Thanks for your attention!

Questions?