SocialWatch: Detection of Online Service Abuse via Large-Scale Social Graphs

Junxian Huang[1] Yinglian Xie[2] Fang Yu[2]
Qifa Ke[2] Martín Abadi[2] Eliot Gillum[3] Z. Morley Mao[1]
[1]University of Michigan [2]Microsoft Research Silicon Valley
[3]Microsoft Corporation

ASIACCS 2013

Arms Race between Attackers and Defenders

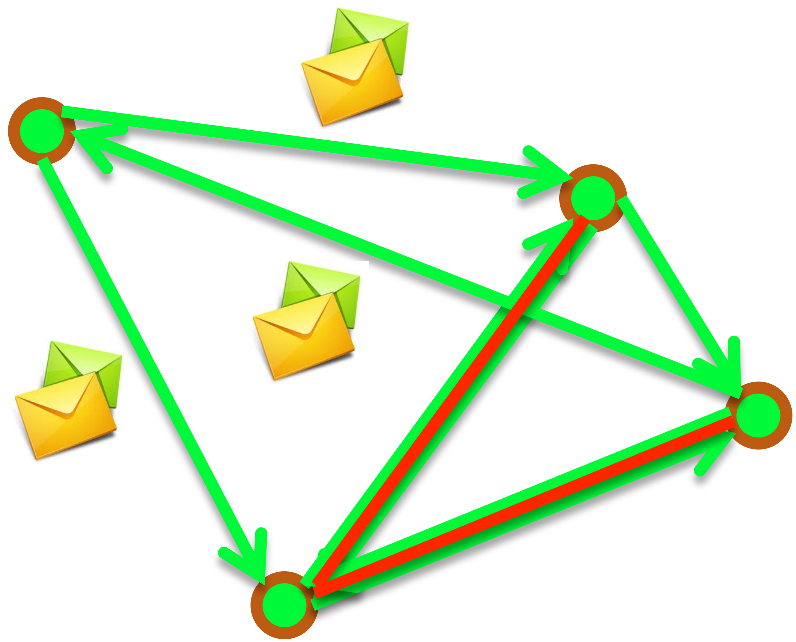
- Malicious accounts in Hotmail
  - Attacker-created accounts
  - Hijacked accounts
  - Attackers are constantly evolving with counter-strategies
- The power of social graph
  - Capture both local and global graph features
  - Hard for attackers to manipulate the overall graph pattern
- Challenges
  - Hijacked accounts have mixed behaviors
  - Incomplete graph – unknown among external accounts
  - Large graph scale requires efficient parallel algorithms

Our Contributions

- Detection methodology – local and global social graph features for detection
- Implementation – demonstrate practicality and scalability for large-scale social graphs
- Evaluation – use a real-world data set with large scale and long duration

Social Graph for Hotmail

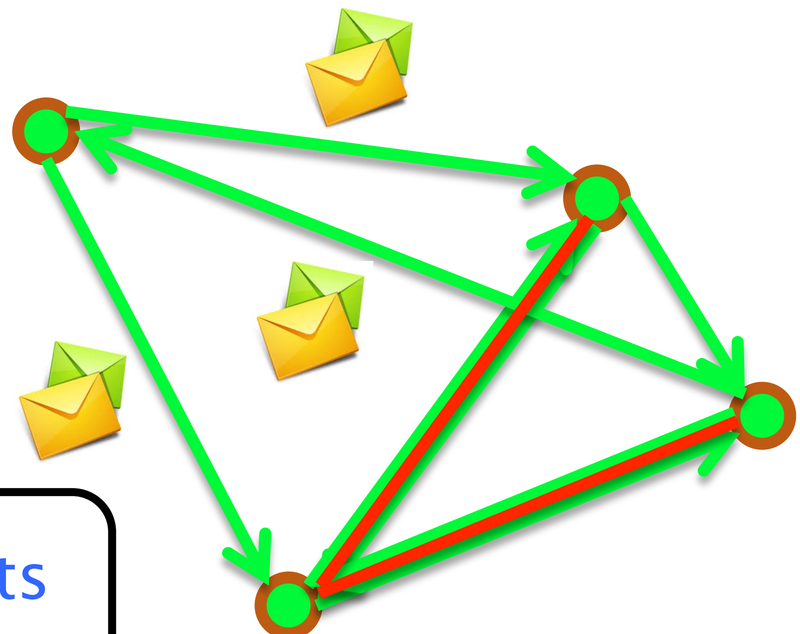
- Vertex
  - Email account
- Edge
  - Directed
    - Send/receive emails
  - Undirected
    - Friendship



Social Graph for Hotmail

Microsoft® Hotmail®

- Vertex
  - Email account (680 million)
- Edge
  - Directed (5.7 billion)
    - Send/receive emails
  - Undirected (440 million)
    - Friendship

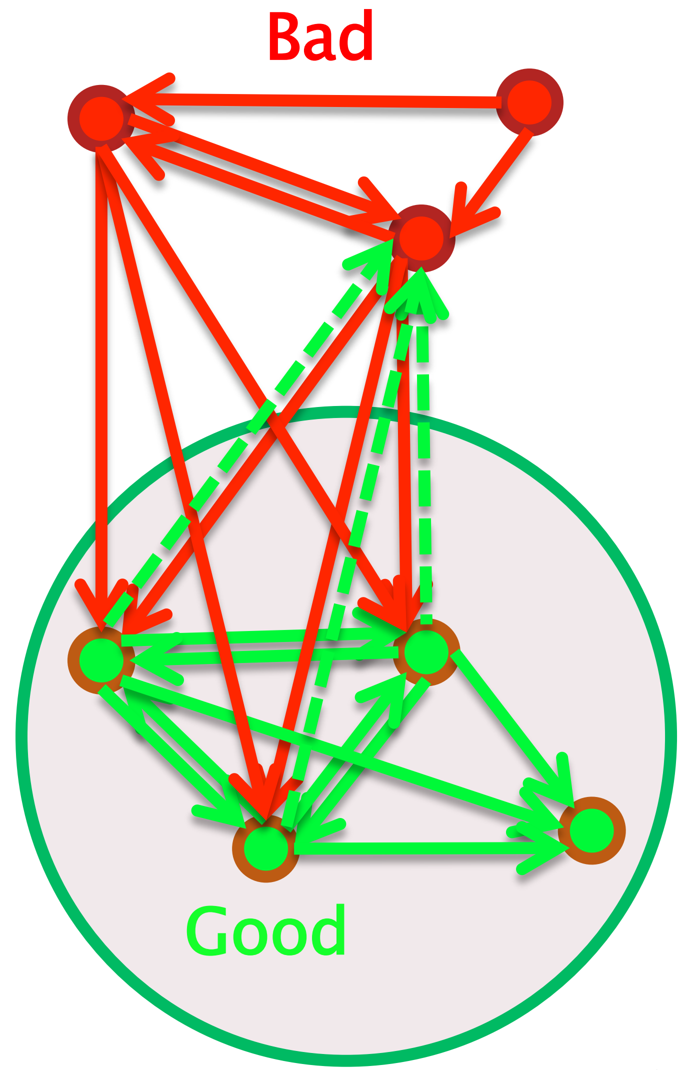


Sampled Hotmail user accounts from 10/2007 to 04/2010

Intuitions in Leveraging Social Graphs

- Good users send emails to other good users
- Sending emails to bad users is suspicious
- Difficult for bad users to enter good users' community

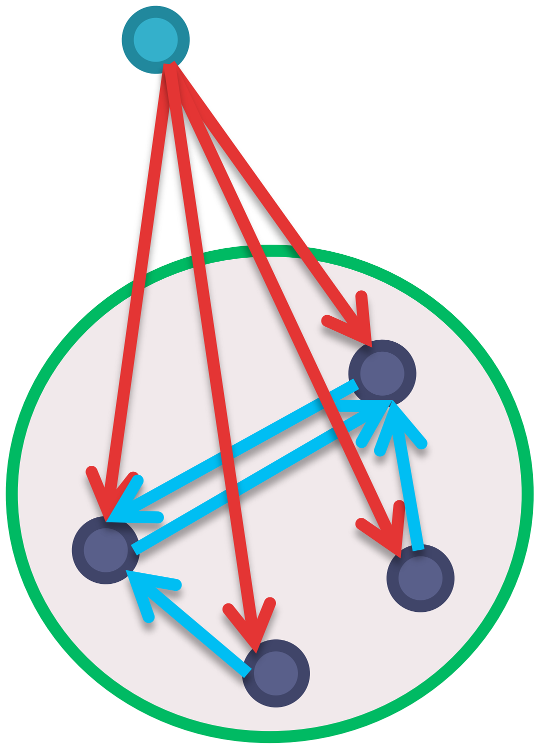
Degree and PageRank based detection



Intuitions in Leveraging Social Graphs

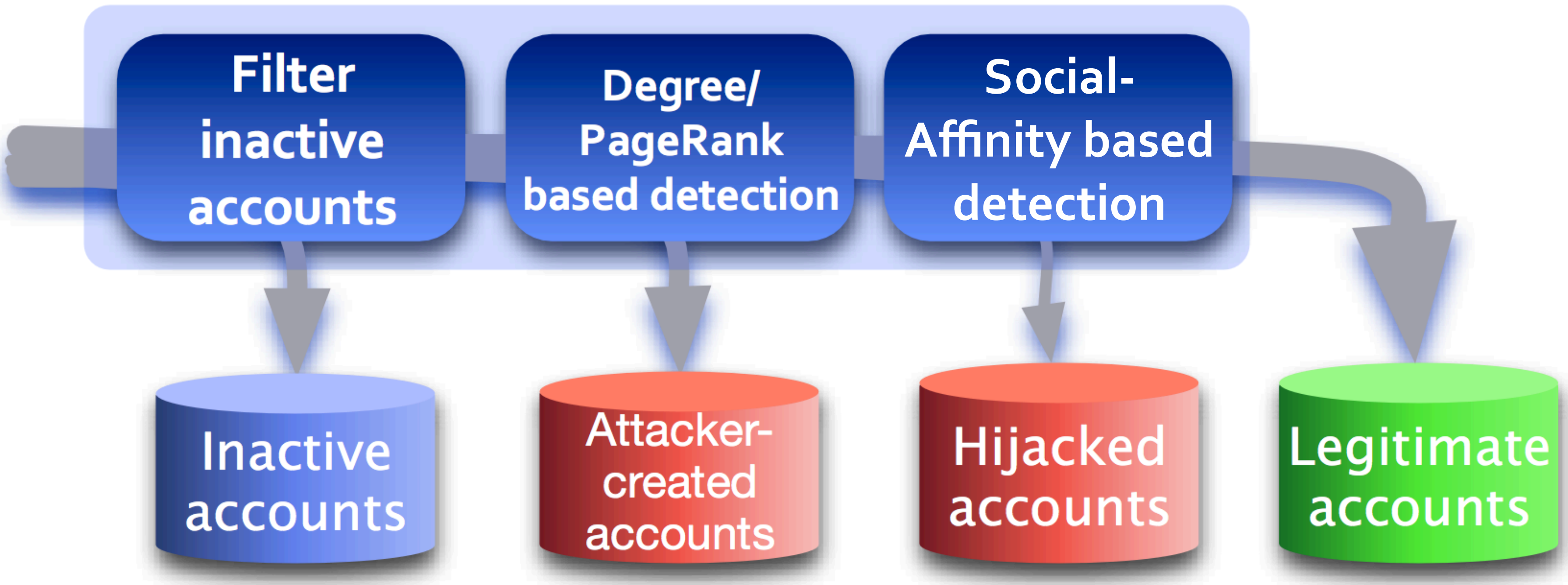
- Recipient sets of good users are more connected than those of bad users

Social-affinity based detection



Recipient set

Design of SocialWatch



Detecting Attacker-created Accounts

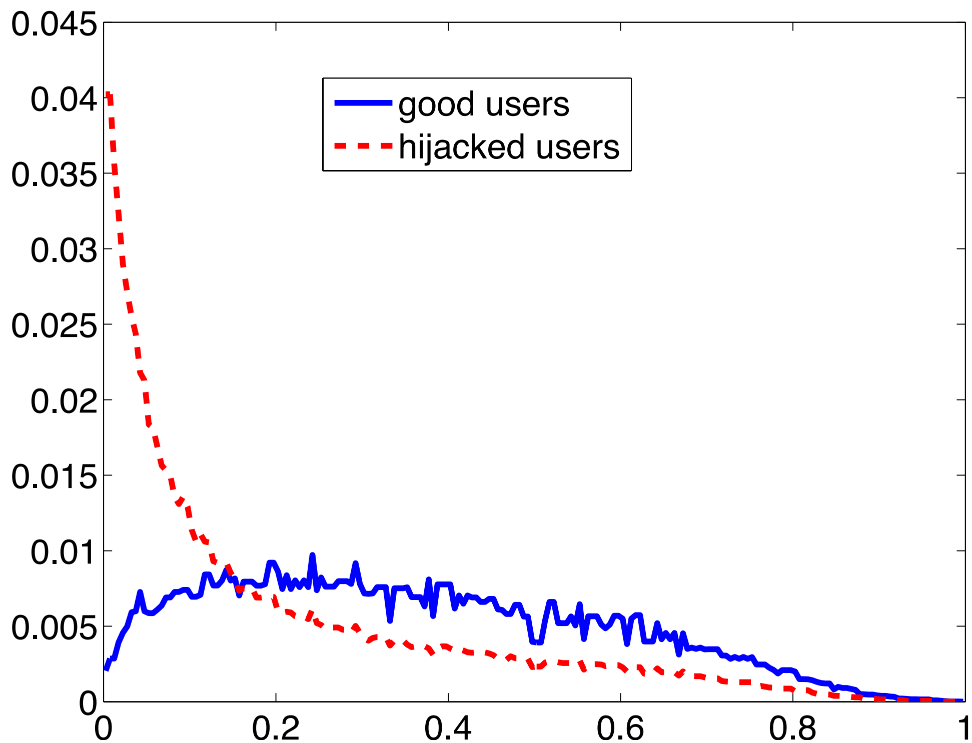
- Social features
  - Degree – a local graph feature that captures the sending/receiving behavior of an account
  - PageRank – a global graph feature that calculates the weight of a node on the overall graph
- Detection methods
  - Identify aggressive spamming accounts with high out degrees and low response rates
  - Identify less aggressive spamming accounts using the badness-goodness PageRank ratio

Computing Goodness/Badness PageRank Score

- Goodness score
  - PageRank value in the directed social graph
- Badness score
  - PageRank value in the reversed directed social graph
- Adjust edge weights based on email exchange patterns
  - Propagate more "goodness" to "good" users and more "badness" to "bad" users

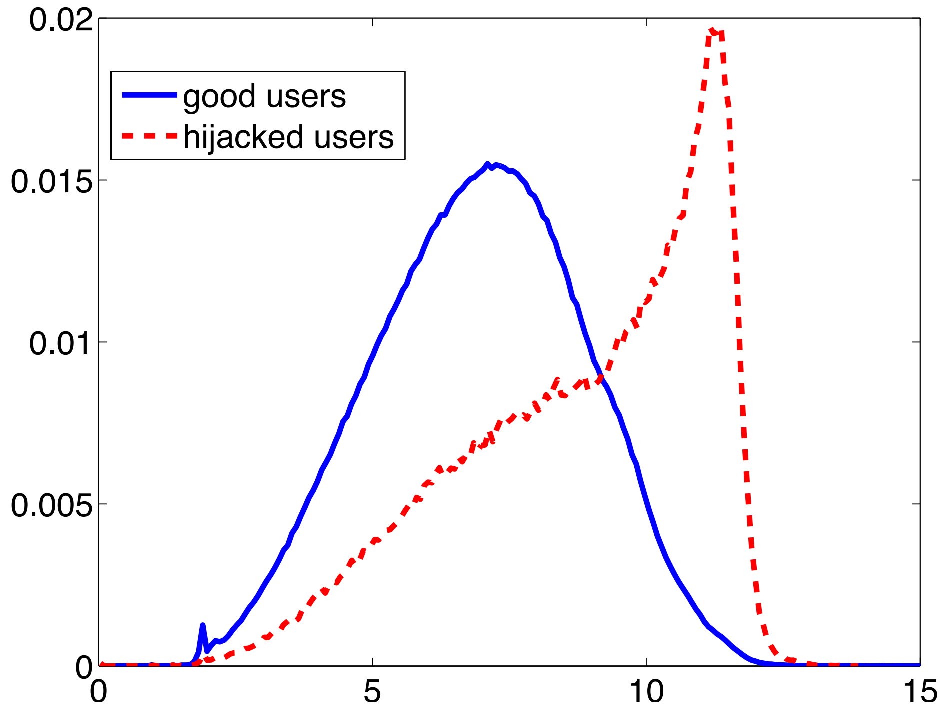
Computing Social-Affinity Features

- Intuition
  - Recipients of legitimate users tend to have more direct connectivity
- Recipient connectivity r
  - The fraction of socially connected recipients



Computing Social-Affinity Features

- Intuition
  - Recipients of legitimate users tend to have closer social distance
- Social distance l
  - The mean of all pairwise social distances between any two users in the recipient set



Detecting Hijacked Accounts

- Detection without known hijacked accounts
  - One-tailed hypothesis testing to detect hijacked accounts
  - Given a significance level, compute a threshold along each feature dimension based on data
  - Classify as hijacked if one of its feature values violates the computed threshold
- Detection with known hijacked accounts
  - Use a Bayesian decision framework to detect additional hijacked accounts using with training data

Implementation and Evaluation

- SocialWatch is implemented using DryadLINQ and processes data in parallel on a 240-machine cluster
- SocialWatch detects 57 million attacker-created accounts, with a 0.8% false detection rate and a 0.6% false negative rate
- At a false detection rate of 2%, SocialWatch identifies 2 million hijacked accounts, 1.2 million were not detected previously

Conclusions

- **SocialWatch** is an online service protection framework, that uses social connectivity features to detect attacker-created accounts and hijacked accounts at a large scale
- SocialWatch is practically deployable and scalable using parallel algorithms

Thank you!

Junxian Huang (hjx@umich.edu)