Sybil Detection and Defense

Yue Duan
Illinois Institute of Technology

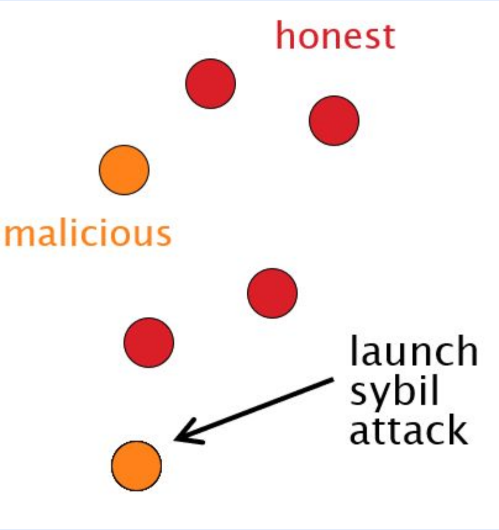
SybilGuard: Defending Against Sybil Attacks via Social Networks

Haifeng Yu, Michael Kaminsky , Phillip B. Gibbons , Abraham Flaxman

SIGCOMM 2006

Background

- Sybil attack
  - Single user pretends many fake/sybil identities
  - Creating multiple accounts from different IP addresses
- Sybil identities can become a large fraction of all identities
  - Out-vote honest users in collaborative tasks

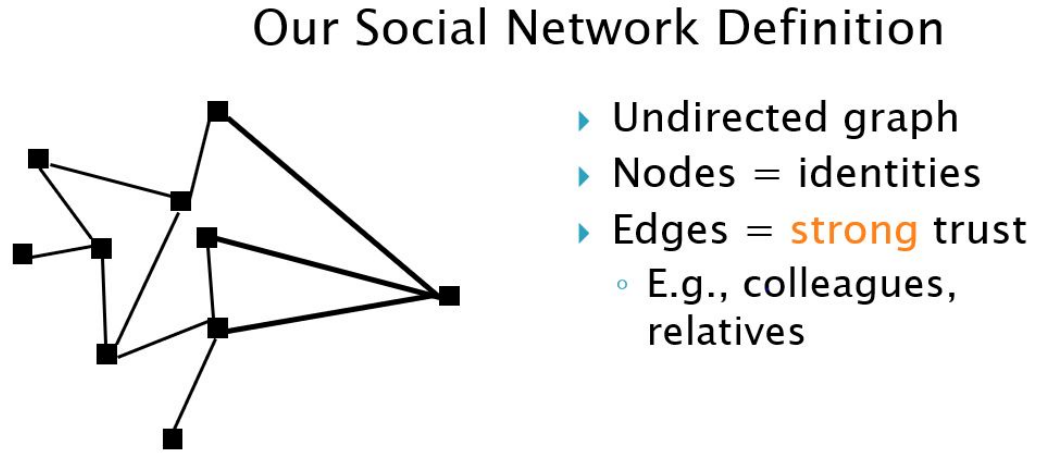


Background

- Defense mechanism
    - Using a trusted central authority
        - Tie identities to actual human beings
    - Not always desirable
        - Can be hard to find such authority
        - Sensitive info may scare away users
        - Potential bottleneck and target of attack
    - Without a trusted central authority
        - Impossible unless using special assumptions [Douceur’02]
        - Resource challenges not sufficient -- adversary can have much more resources than typical user

SybilGuard

- Main Idea: Use a social network as the “central authority”
- A node trusts its neighbors
- Each node learns about the network from its neighbors

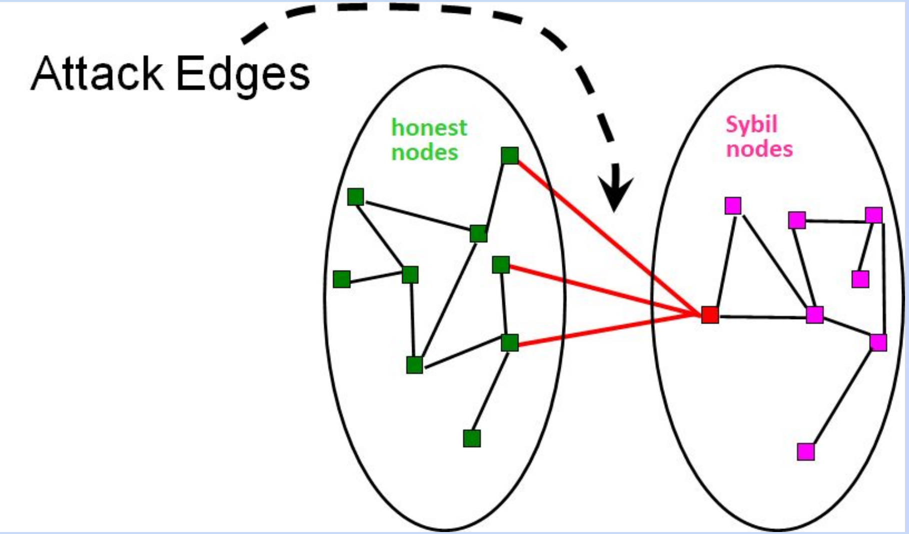


Our Social Network Definition

- Undirected graph
- Nodes = identities
- Edges = strong trust
  - E.g., colleagues, relatives

Sybil Nodes and Attack Edges

- Edges to honest nodes are “human established”
- Attack edges are difficult for Sybil nodes to create
- Attack edges are **rare**
    - To subvert system an attacker must compromise many honest nodes

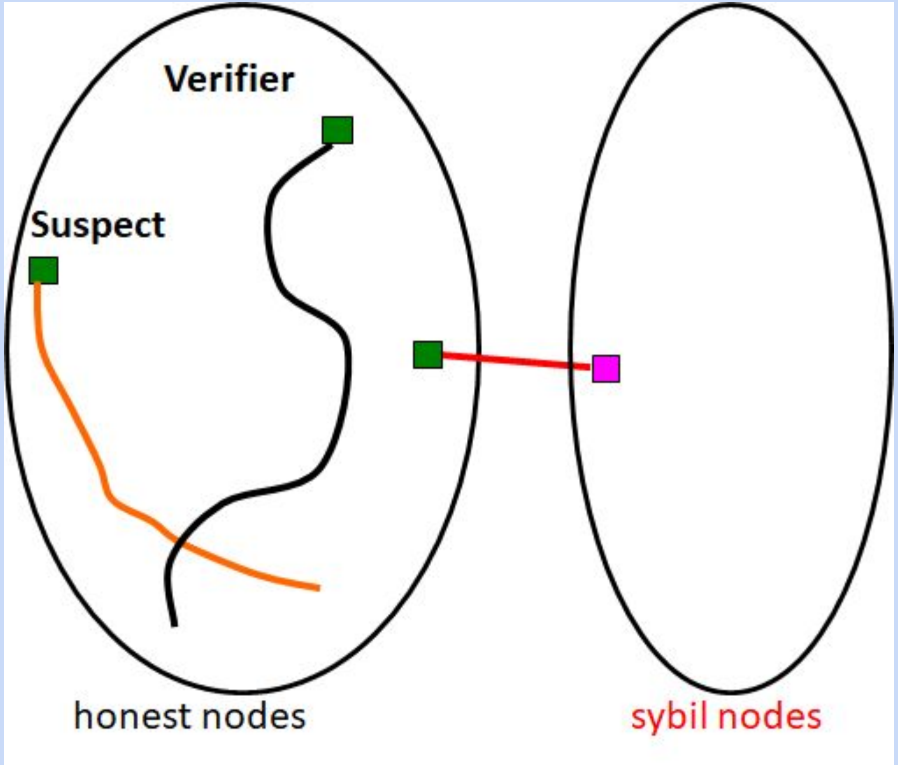


SybilGuard

- A social network exists containing honest nodes and Sybil nodes
- Honest nodes provide a service to or receive a service from nodes that they "accept"
- Ideally, only honest nodes are accepted
- With high probability an honest nodes
  - Accepts most honest nodes
  - Is accepted by most honest nodes
  - Accepts at most a bounded number of Sybil nodes

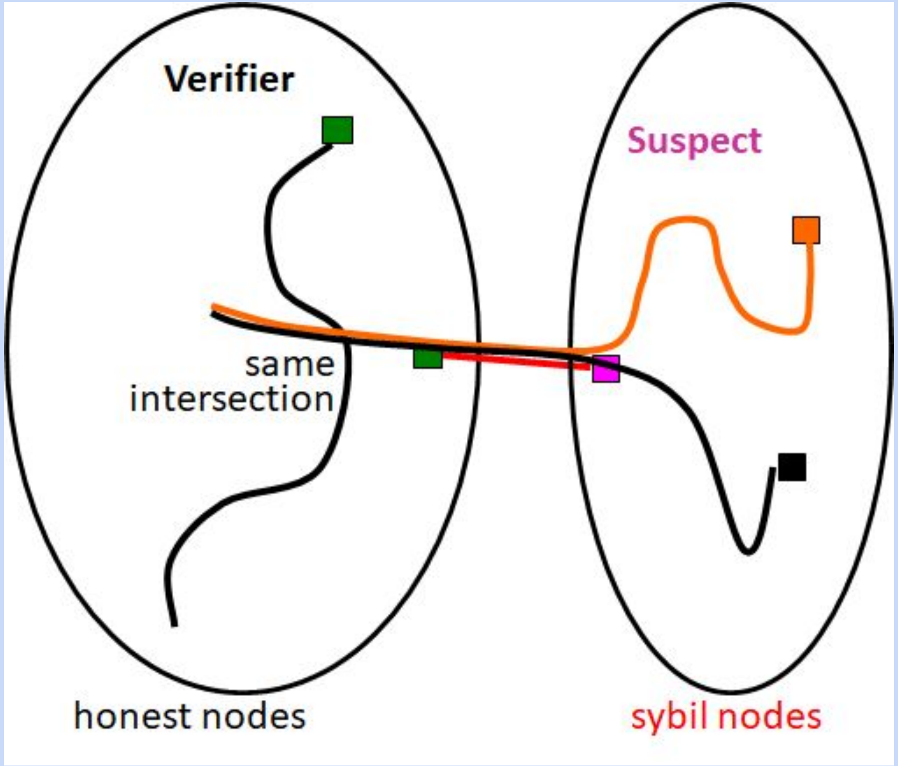
Random Route Intersection

- Random walk
  - Each node finds all the length w random routes that start at it
  - Honest node V accepts node S if most of V’s random routes intersect a random route of S
- With high probability
  - verifier’s route stays within honest region
  - routes from two honest nodes intersect



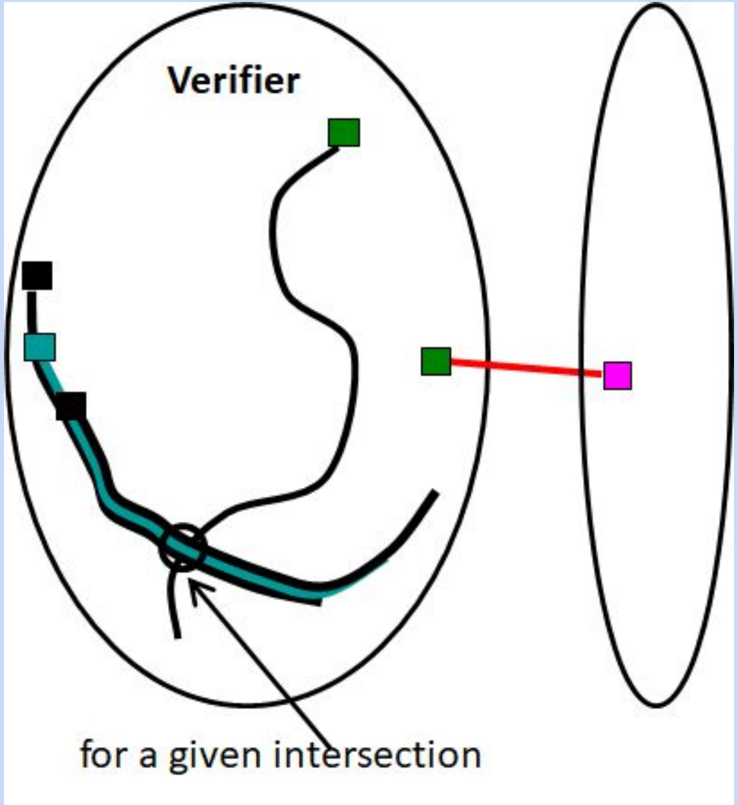
Random Route Intersection

- Each attack edge gives one intersection
- Intersection points are SybilGuard’s equivalence sets



Random Route Intersection

- Verifier accepts at most w nodes per intersection



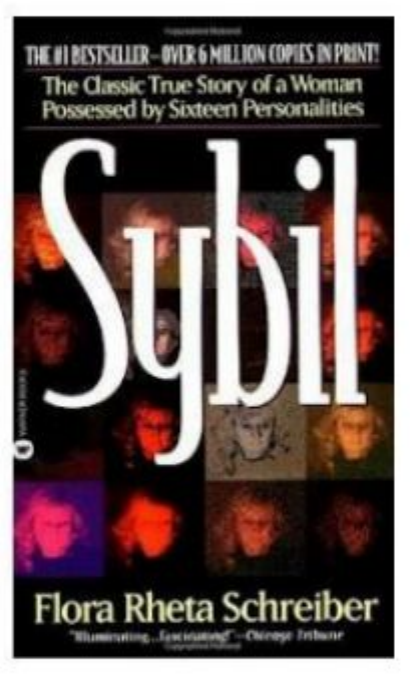
Uncovering Social Network Sybils in the Wild

Zhi Yang, Christo Wilson, Xiao Wang, Tingting Gao, Ben Y. Zhao, Yafei Dai

Peking University, UC Santa Barbara

IMC 2011

Sybil, fake account


THE #1 BESTSELLER – OVER 6 MILLION COPIES IN PRINT!
The Classic True Story of a Woman Possessed by Sixteen Personalities
Sybil
Flora Rheta Schreiber
"Illuminating...fascinating" —Chicago Tribune

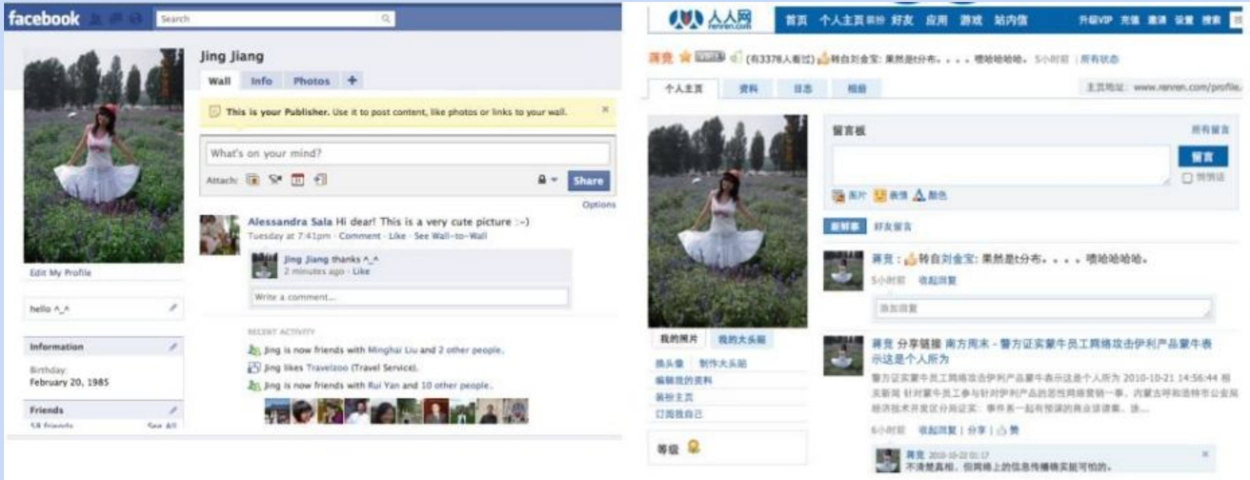
Sybil, Noun

: a book of which content is a case study of a woman diagnosed with multiple personality disorder

“a fake account that attempts to create many friendships with honest users”

Target: Renren

- Renren: oldest and largest OSN in China



This is your Publisher. Use it to post content, like photos or links to your wall.

What's on your mind?

Alessandra Sala Hi dear! This is a very cute picture :-)
Tuesday at 7:41pm · Comment · Like · See Wall-to-Wall

Jing Jiang thanks ^_^
2 minutes ago · Like

Write a comment...

Jing is now friends with Minghai Liu and 2 other people.
Jing likes Travelzoo (Travel Service).
Jing is now friends with Rui Yan and 10 other people.

蒋竞 (有3378人看过) 转自刘金宝: 果然是t分布。。。。嘿哈哈哈哈。 5小时前 | 所有状态

蒋竞：转自刘金宝: 果然是t分布。。。。嘿哈哈哈哈。
5小时前 收起回复

蒋竞 分享链接 南方周末 - 警方证实蒙牛员工网络攻击伊利产品蒙牛表示这是个人所为
警方证实蒙牛员工网络攻击伊利产品蒙牛表示这是个人所为 2010-10-21 14:56:44 相关新闻 针对蒙牛员工参与针对伊利产品的恶性网络营销一事，内蒙古呼和浩特市公安局经济技术开发区分局证实：事件系一起有预谋的商业诽谤案。涉...
6小时前 收起回复 | 分享 | 赞

蒋竞 2010-10-22 01:17
不清楚真相，但网络上的信息传播确实挺可怕的。

Previous detector on Renren

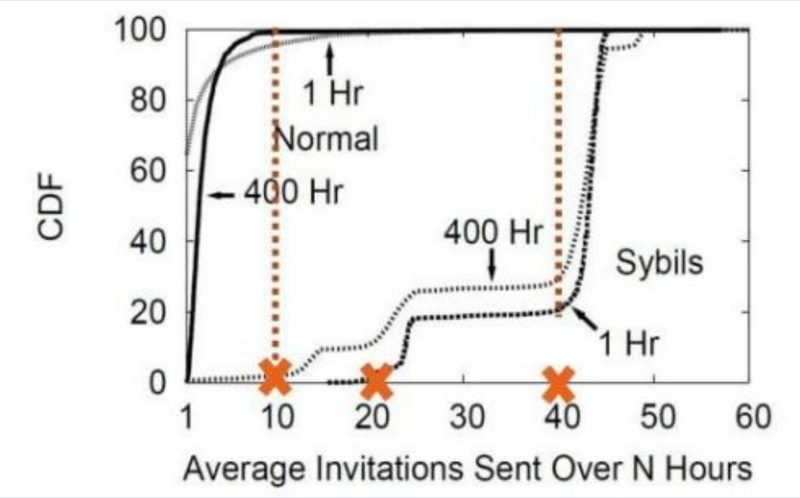
- Using orthogonal techniques to find sybil accounts
  - spamming & scanning content for suspect keywords and blacklisted URLs
  - crowdsourced account flagging
- Detect results
  - 560 sybils banned as of Aug 2010
- Limitations:
  - ad-hoc
  - require human effort
  - operate after posing spam content

Improved Detector

- Developed improved Sybil detector for Renren
    - Analyze ground-truth data on existing sybils
        - find behavioral attributes to identify sybil accounts
        - examine a wide range of attributes
        - find four potential identifiers

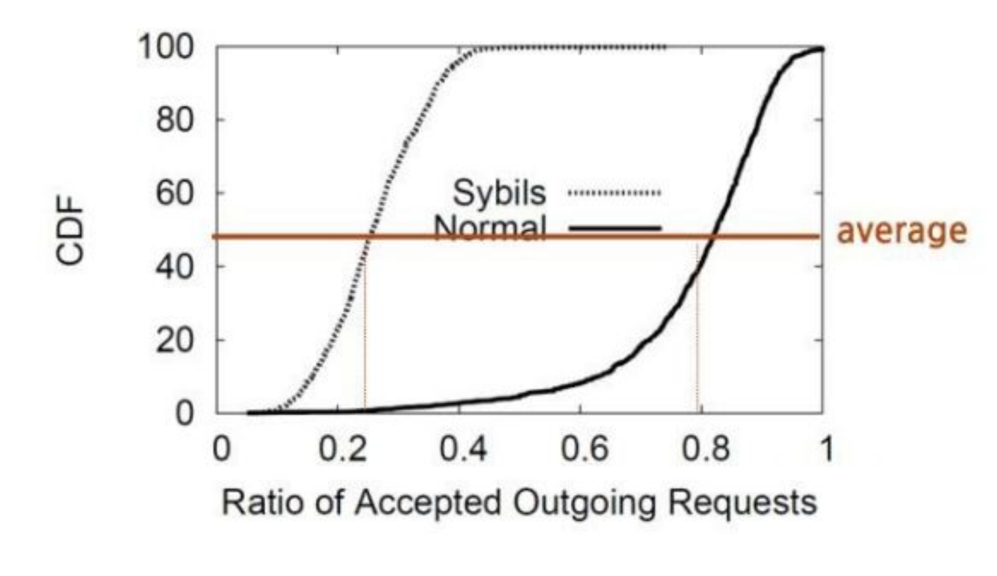
Four Reliable Sybil Indicators

- Friend request frequency (invitation frequency)
  - the number of friend requests a user has sent within a fixed time period



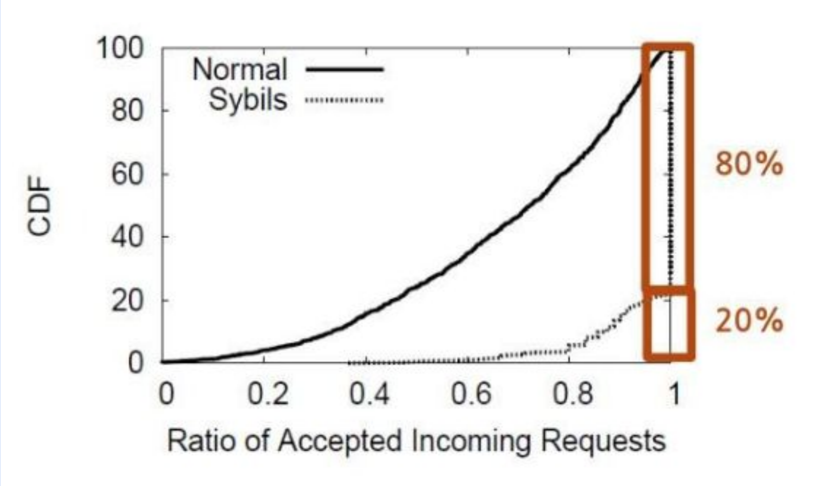
Four Reliable Sybil Indicators

- Outgoing friend request accepted
    - requests confirmed by the recipient



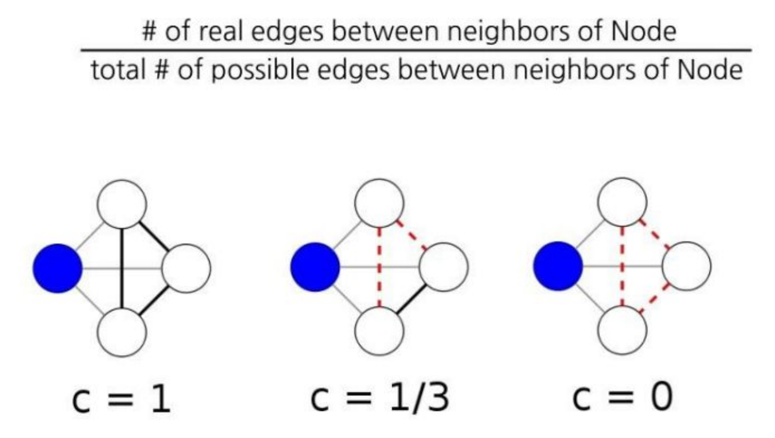
Four Reliable Sybil Indicators

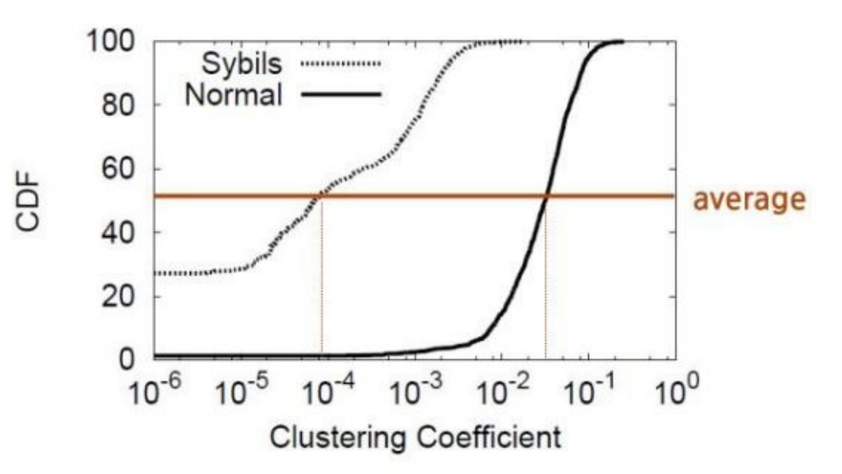
- Incoming friend request accepted
  - The fraction of incoming friend requests accepted



Four Reliable Sybil Indicators

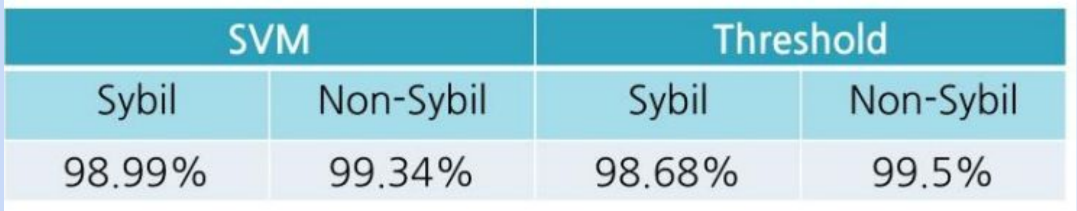
- Clustering coefficient
  - a graph metric that measures the mutual connectivity of a user's friends


# of real edges between neighbors of Node
total # of possible edges between neighbors of Node
c = 1
c = 1/3
c = 0



Verify Sybil Detector

- Evaluate threshold and SVM detectors
  - dataset: 1000 normal user and 1000 sybils
  - similar accuracy for both



SVM		Threshold	
Sybil	Non-Sybil	Sybil	Non-Sybil
98.99%	99.34%	98.68%	99.5%

  - deployed threshold, less CPU intensive, real-time
  - adaptive feedback scheme is used to dynamically tune threshold parameters

Detection Results

- Detect 100K sybils in the first six months (aug 2010 - feb 2011)
    - vast majority (67%) are spammers
- Low false positive rate
    - use customer complaint rate as signal
    - complaints evaluated by humans
    - 25 **real** complaints per 3000 bans (<1%)

spammers attempted to recover banned sybils by complaining to renren.