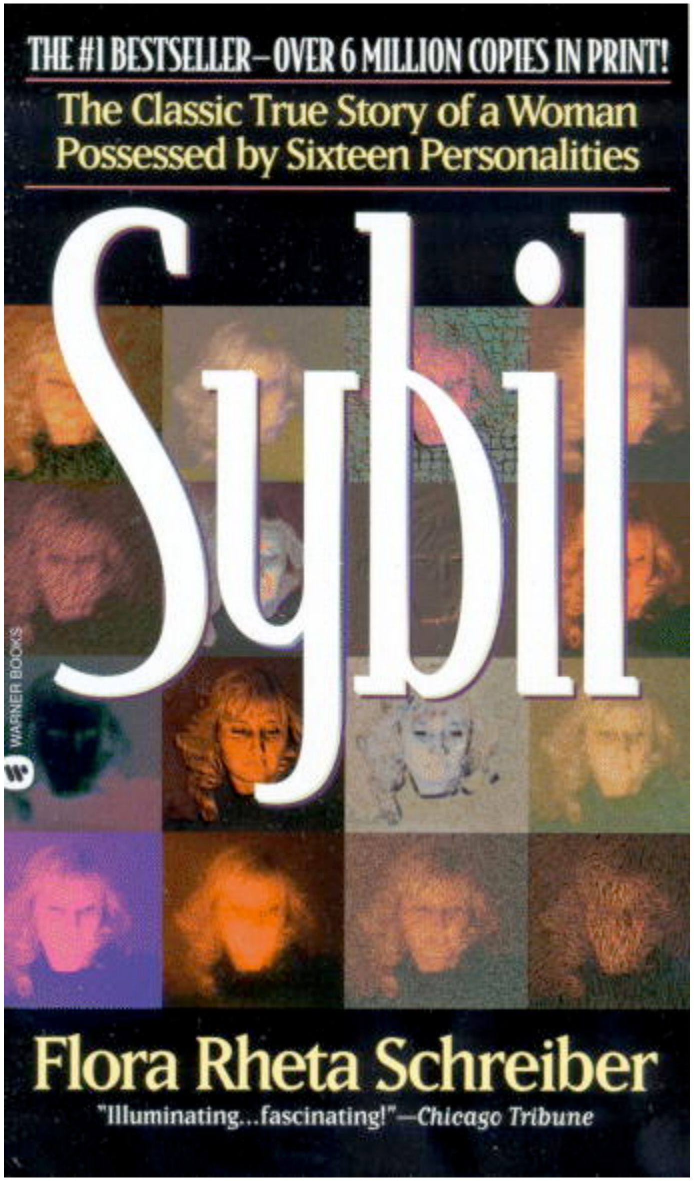
Sybils

EEECE 571B “Computer Security”

Konstantin Beznosov


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

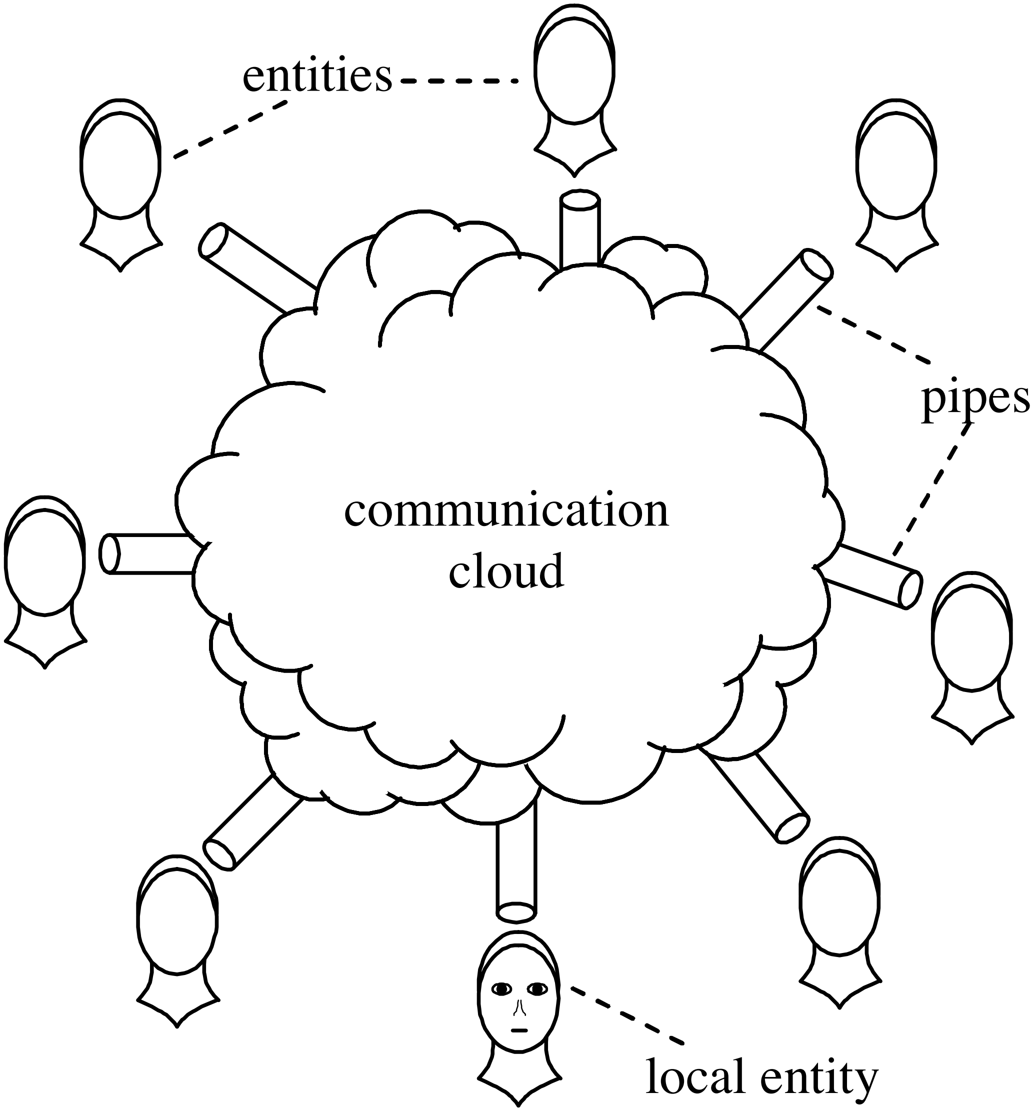
dissociative identity disorder (multiple personality disorder)

the risk of the Sybil attack in a nutshell

- inherent vulnerability exploited in any Sybil Attack: it is always possible for an entity to present multiple distinct identities.
- redundancy lets distributed systems compensate for faulty nodes
  - Ex: Store data on multiple nodes
- the Sybil Attack undermines redundancy
  - need a central authority to determine which nodes are honest

reasoning model

1. An entity can send a message through its pipe, broadcasting it to all other entities.
2. The message will be received by all entities within a bounded interval of time.
3. Message delivery is guaranteed, but there is no assurance that all entities will hear messages in the same order.
4. Entities can establish virtual point-to-point communication paths that are private and authenticated.



adopted from [1]

entity vs. identity

- An identity is an abstract representation that persists across multiple communication events.
- Each <u>entity</u> *e* attempts to present an <u>identity</u> *i* to other entities in the system.
- If *e* successfully presents identity *i* to *l*, we say that *l* <u>accepts</u> identity *i*.
- Each <u>correct entity</u> *c* will attempt to present one legitimate identity.
- Each <u>faulty entity</u> *f* may attempt to present a legitimate identity and one or more counterfeit identities.
- The system should accept all legitimate identities but no counterfeit entities.
- An entity has three potential sources of information about other entities: a trusted agency, itself, or other (untrusted) entities.

direct validation of entities (from [1])

- A faulty entity can counterfeit a constant number of multiple identities.
  - Lemma 1: “If *p* is the ratio of the resources of a faulty entity to the resources of a minimally capable entity, then *f* can present *g*=floor(*p*) distinct identities to local entity *L*”
  - lower bound -> upper bound
- Each correct entity must simultaneously validate all the identities it is presented; otherwise, a faulty entity can counterfeit an unbounded number of identities.
  - Lemma 2: “If a local entity *L* accepts entities that are not validated simultaneously, then a single faulty entity *f* can present an arbitrarily large number of distinct identities to L”

for indirect validation

- A sufficiently large set of faulty entities can counterfeit an unbounded number of identities.
  - Lemma 3: “If local entity L accepts any identity vouched for by q accepted identities, then a set F of faulty entities can present an arbitrarily large number of distinct to L if either |F|>=q, or the collective resources available to F at least equals q+|F| minimally capable entities”
- All entities in the system must perform their identity validations concurrently; otherwise, a faulty entity can counterfeit a constant number of multiple identities.
  - Lemma 4: “If the correct entities in set *C* do not coordinate time intervals during which they accept identities, and if local entity *L* accepts any identity vouched for by *q* accepted identities, then even a minimally capable faulty entity *f* can present *g=floor(*|*C*|*/q)* distinct identities to *L*.”

so what?

if there is no identification authority

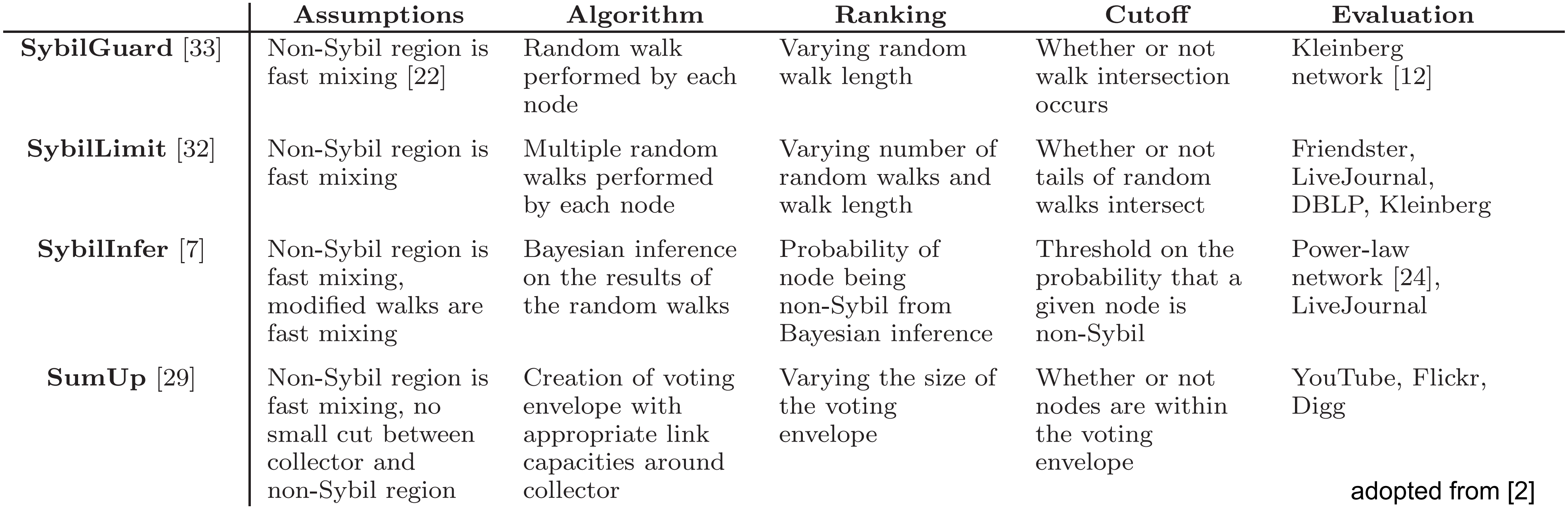
- one has to assume that an attacker’s resources are limited
- resource-demanding challenges to validate identities
- conditions:
  1. all entities operate under nearly identical resource constraints
  2. all presented identities are validated simultaneously by all entities
  3. when accepting identities that are not directly validated, the required number of vouchers exceeds the number of system-wide failures.

are these conditions justifiable as assumptions and practically realizable?

history of countering online Sybils

- computational games and CAPTCHAs to increase the cost of creating identities
- detection of Sybils
  - based on trust and reputation (Advogato, Appleseed, SybilProof)
    - vulnerable to whitewashing attacks, where attackers initially behave honestly
  - community detection
    - Sybilguard, Sybillimit, Sybilinfer, Tran et al., SumUp, Whanau
    - assumption: an attacker cannot establish an arbitrarily large number of social connections to non-Sybil nodes ==> Sybil nodes are poorly connected to the rest of the network
  - use of account-related statistics
    - outgoing request accepted ratio, invitation frequency, clustering coefficient, etc.

social network-based Sybil defenses



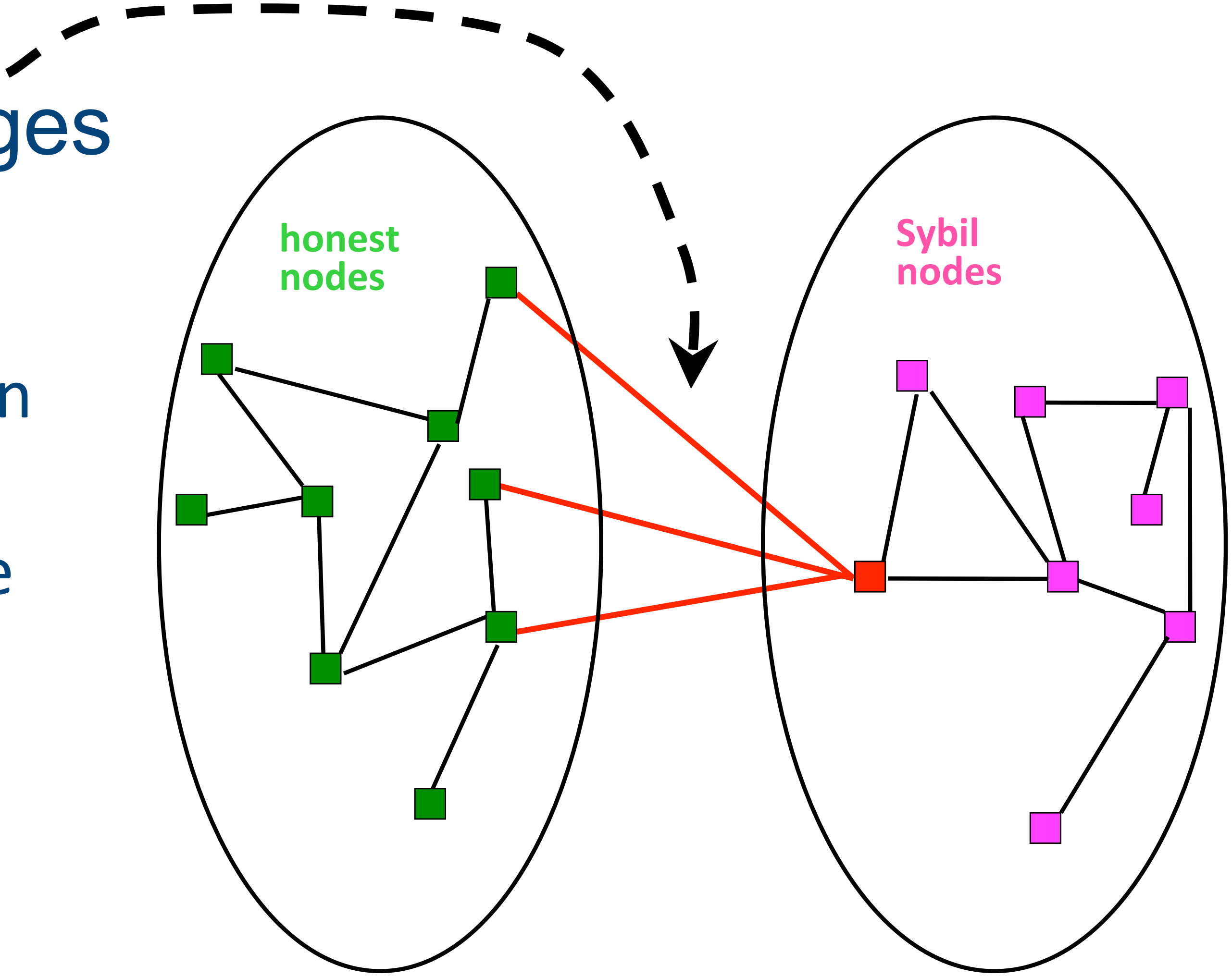
	Assumptions	Algorithm	Ranking	Cutoff	Evaluation
SybilGuard [33]	Non-Sybil region is fast mixing [22]	Random walk performed by each node	Varying random walk length	Whether or not walk intersection occurs	Kleinberg network [12]
SybilLimit [32]	Non-Sybil region is fast mixing	Multiple random walks performed by each node	Varying number of random walks and walk length	Whether or not tails of random walks intersect	Friendster, LiveJournal, DBLP, Kleinberg
SybilInfer [7]	Non-Sybil region is fast mixing, modified walks are fast mixing	Bayesian inference on the results of the random walks	Probability of node being non-Sybil from Bayesian inference	Threshold on the probability that a given node is non-Sybil	Power-law network [24], LiveJournal
SumUp [29]	Non-Sybil region is fast mixing, no small cut between collector and non-Sybil region	Creation of voting envelope with appropriate link capacities around collector	Varying the size of the voting envelope	Whether or not nodes are within the voting envelope	YouTube, Flickr, Digg

adopted from [2]

fast-mixing: a random walk of length O(log N) reaches a stationary distribution of nodes

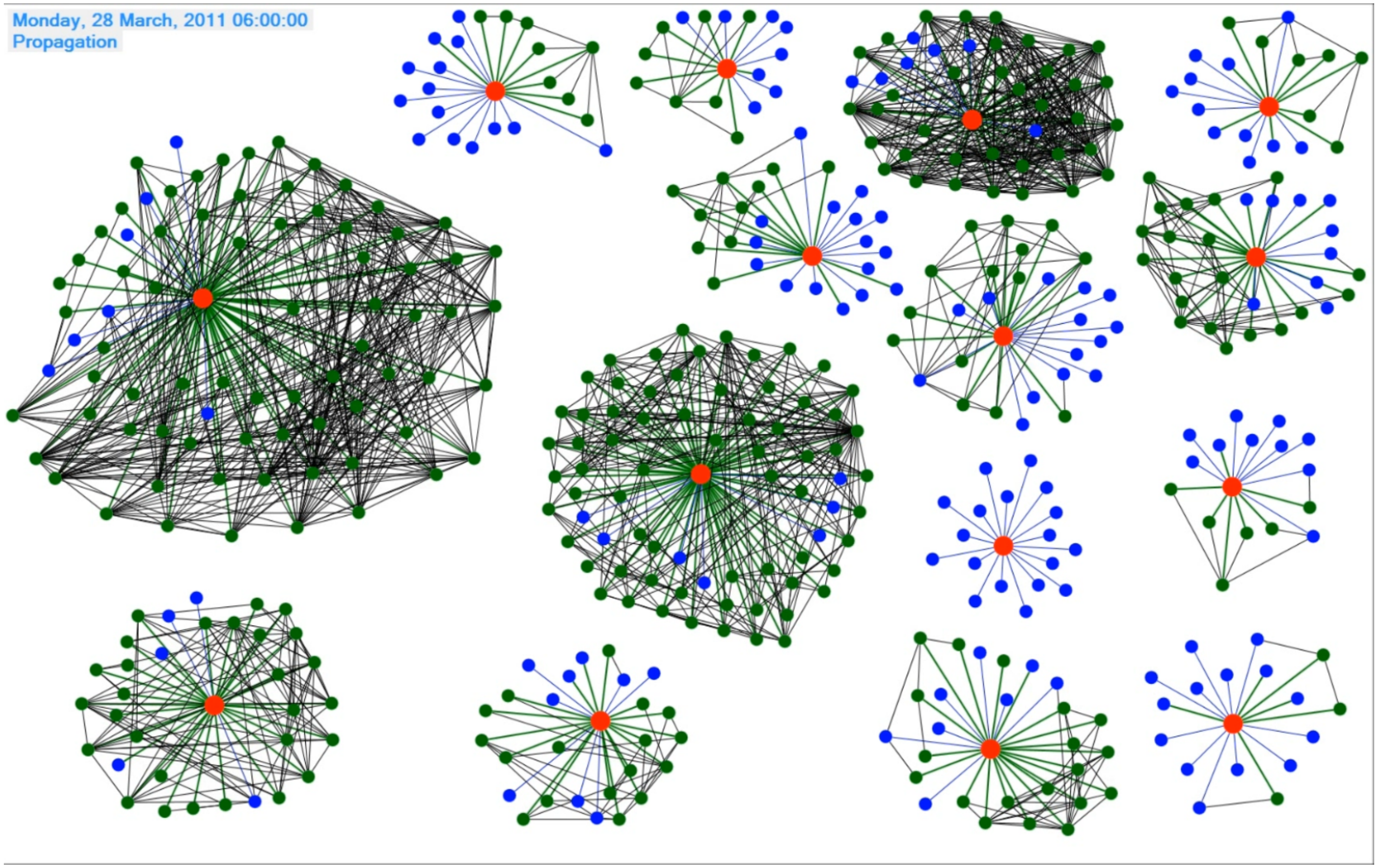
Sybil nodes and attack edges

Attack Edges



- Edges to honest nodes are “human established”
- Attack edges are difficult for Sybil nodes to create

Sybil nodes can blend with the honest ones



8,570 requests sent, 3,055 accepted

Latent Community Model for Detecting Sybils

Z. Cai, C. Jermaine, “The Latent Community Model for Detecting Sybils in Social Networks,” NDSS ’12

assumptions

1. A special set of size s of the graph’s nodes is known to be benevolent; they are called the “seeds”.
2. Nodes in the same community are either uniformly malicious or uniformly benign.
   - nodes within communities are (by definition) connected with a uniform density
   - “... it seems unlikely that a set of malicious nodes would be able to so thoroughly integrate themselves into a community of benign nodes that there is no real difference in the connection density between the benign nodes in the community and the attackers ...”
   - “Even if such an integration did occur, those benign nodes would be so thoroughly compromised that labeling them as attackers would not be an egregious error.”

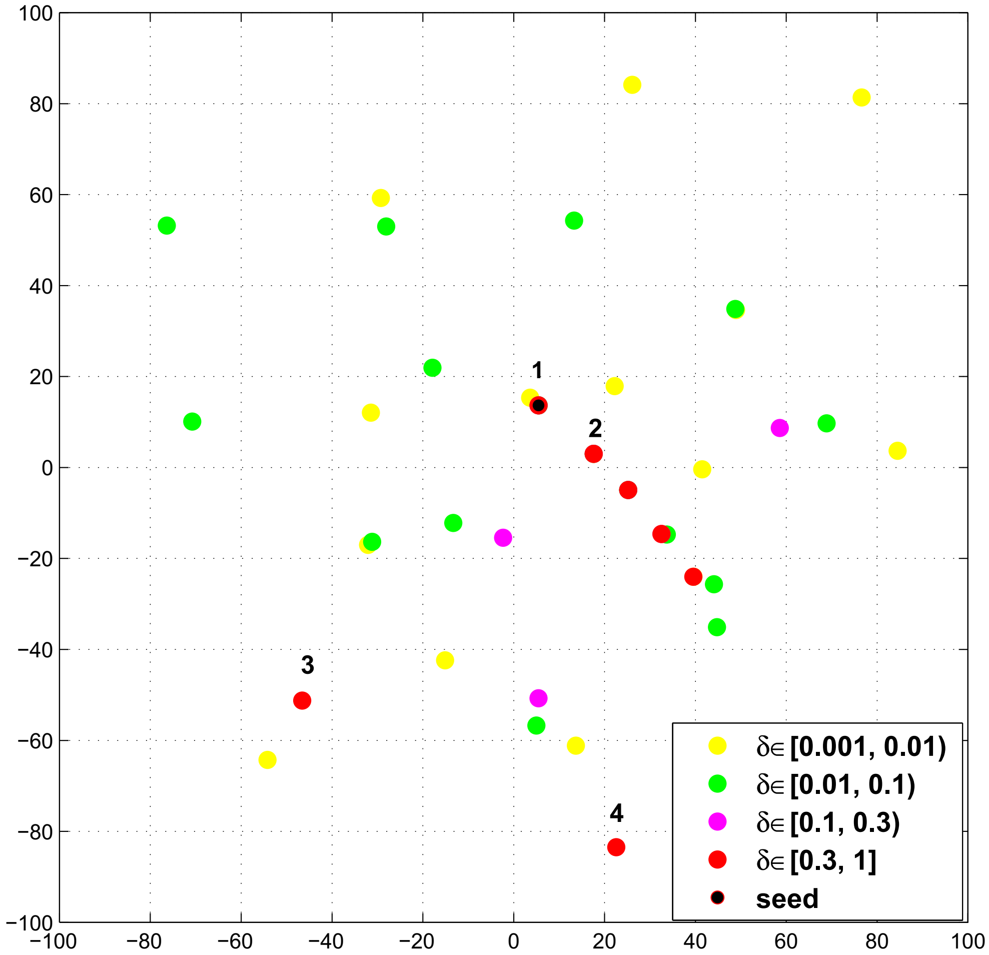
intuition of the LC model

- learned communities positioned in a latent (Euclidean space)
- communities near each other tend to have a large number of connections
- if the attack communities are attached to the “good” portion of the network in a way that is inconsistent with other communities, they will tend to be pushed to the “outside” of the the latent space.

LC model

- community: set of nodes with (relatively) dense interconnections
- community is associated with a latent position in a multidimensional Euclidean space
- communities that are close have many links between them
  - far apart communities have few links
- Gaussian distribution positions the benign communities close to the center of the space
- spherical distribution of attackers that surrounds the Gaussian distribution

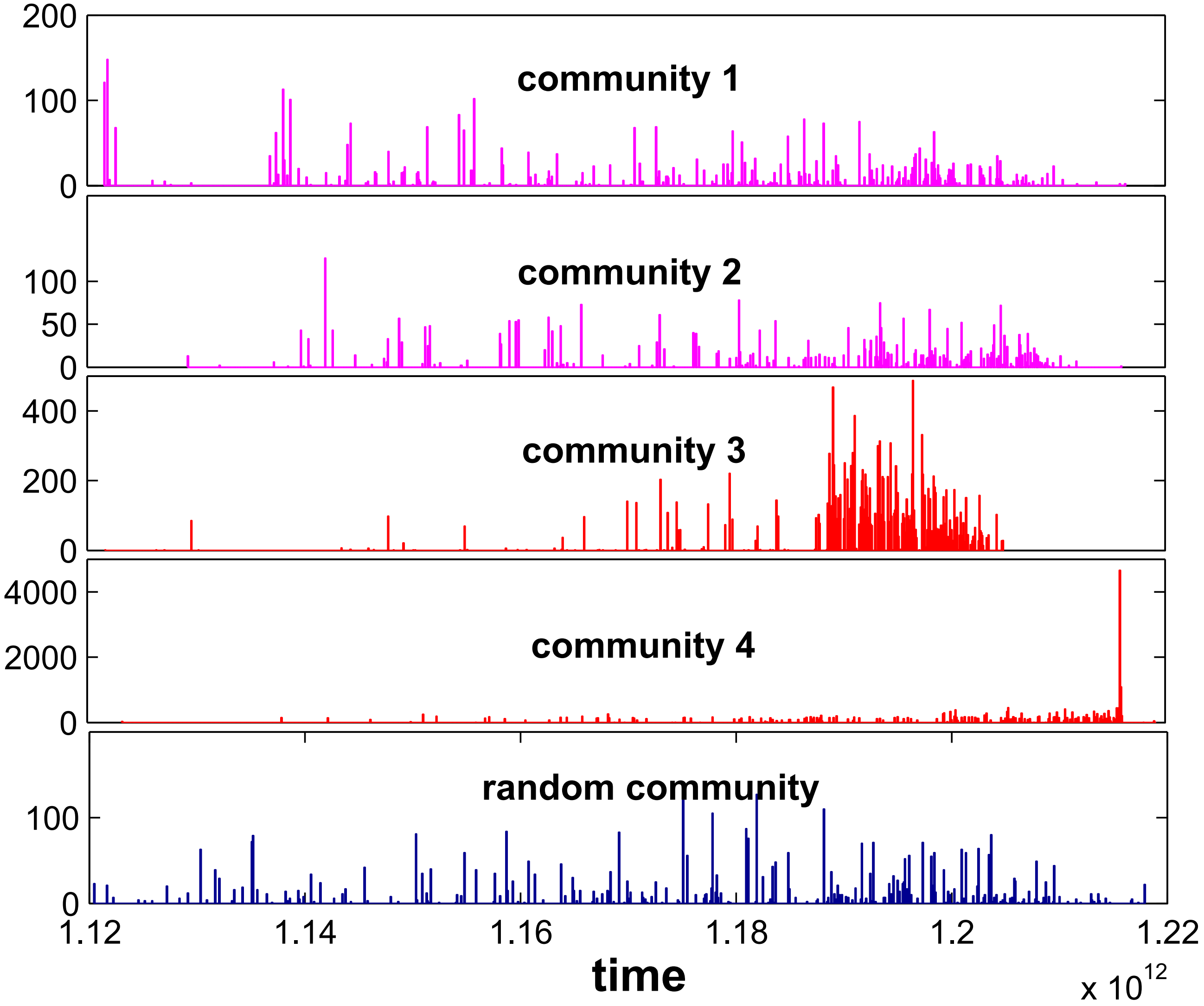
example: Digg



- 594, 426 nodes, 4, 070, 026 undirected edges

adopted from [3]

creation time of edges



adopted from [3]

credits

These slides incorporate parts of the following:

1. Douceur, J.R. “The Sybil attack” in First International Workshop Peer-to-Peer Systems, IPTPS, 2002 Cambridge, MA, USA, March 7-8, 2002, pp. 251-260.
2. Bimal Viswanath, Ansley Post, Krishna P. Gummadi, and Alan Mislove, “An analysis of social network-based Sybil defenses,” In Proceedings of the ACM SIGCOMM 2010 conference on SIGCOMM (SIGCOMM ’10). ACM, New York, NY, USA, 363-374.
3. Z. Cai, C. Jermaine, “The Latent Community Model for Detecting Sybils in Social Networks,” NDSS ’12
4. “DSybil: Optimal Sybil-Resistance for Recommendation Systems” (Oakland ’09)
5. Bruschi, D.; Cavallaro, L.; Lanzi, A., “An Efficient Technique for Preventing Mimicry and Impossible Paths Execution Attacks,” Performance, Computing, and Communications Conference, 2007. IPCCC 2007. IEEE Internationa , vol., no., pp.418-425, 11-13 April 2007
6. Zhi Yang, Christo Wilson, Xiao Wang, Tingting Gao, Ben Y. Zhao, and Yafei Dai. 2011. Uncovering social network sybils in the wild. In Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference (IMC ’11). ACM, New York, NY, USA, 259-268.
7. Haifeng Yu, Michael Kaminsky, Phillip B. Gibbons, and Abraham Flaxman, “SybilGuard: defending against sybil attacks via social networks” SIGCOMM Comput. Commun. Rev. 36, 4 (August 2006), 267-278. DOI=10.1145/1151659.1159945