ISSN: 2321-7782 (Online)
Volume 2, Issue 12, December 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Analyze and Detect Sybil Nodes in Online Community Networks Using Structure Based

Dr. K. Swathi[1]
Professor and Head
Department of Computer science &Engineering
Cauvery College of Engineering & Technology
Trichy (TN) - India

P. Nidhya[2]
Scholar
Department of Computer science &Engineering
Cauvery College of Engineering & Technology
Trichy (TN) – India

R. Vijayanathan[3]
Senior Librarian and Head,
Department of Library and Information Science
Cauvery College of Engineering & Technology
Trichy (TN) - India

Abstract: ***SybilBelief, a semi-supervised learning framework, to detect Sybil nodes. SybilBelief it takes a of the nodes in the social network of the system and optionally, a small set of known Sybils as input. Then, SybilBelief propagates the label information from the known benign and/or Sybil nodes to the remaining nodes in the system. In SybilBelief we process both synthetic and real-world social network topologies. SybilBelief is able to accurately identify Sybil nodes with low false positive rates and low false negative rates. SybilBelief is resilient to noise in our prior knowledge about known benign and Sybil nodes. A number of IP traceback approaches have been suggested to identify attackers and there are two major methods for IP traceback, the probabilistic packet marking (PPM) and the deterministic packet marking (DPM) Both of these strategies require routers to inject marks into individual packets. So newly introduced effective and efficient IP traceback scheme against DoS attacks based on entropy variations. Trackback mechanisms identifying the number of zombies in large scale network and all so give the authentication for blocked users. It works as an independent software module with current routing software. This makes it a feasible and easy to be implemented solution for the current Internet.***

Keywords: Sybil detection, semi-supervised learning, Markov random fields, belief propagation

I. INTRODUCTION

Sybil attacks, where a single entity emulates the behavior of multiple users, form a fundamental threat to the security of distributed systems. Example systems include peer-to-peer networks, email, reputation systems, and online social networks. Sybil accounts in online social networks are used for criminal activities such as spreading spam or malware, stealing other users' private information, and manipulating web search results via "+1" or "like" clicks. Traditionally, Sybil defenses require users to present trusted identities issued by certification authorities. However, such approaches violate the open nature that underlies the success of these distributed systems. Recently, there has been a growing interest in leveraging social networks to mitigate Sybil attacks. These schemes are based on the observation that, although an attacker can create arbitrary Sybil users and social connections among themselves, he or she can only establish a limited number of social connections to benign users. As a result, Sybil users tend to form a community structure among them, which enables a large number of Sybil users to integrate into the system. Note that it is crucial to obtain social connections that represent trust relationships between users; otherwise the structure-based Sybil detection mechanisms have limited detection accuracy a semi-supervised learning problem, where the goal is to propagate reputations from a small set of known benign and/or Sybil users to other users along the social connections between them. More specifically, we first associate a binary random variable with each user in the system; such random variable represents the label (i.e., benign or Sybil) of the user. Second, we model the social network between users in

the system as a pair wise Markov Random Field, which defines a joint probability distribution for these binary random variables. Third, given a set of known benign and/or Sybil users, we infer the posterior probability of a user being benign, which is treated as the reputation of the user. For efficient inference of the posterior probability, we couple our framework with Loopy Belief Propagation, an iterative algorithm for inference on probabilistic graphical models.

II. Related Work

A structure-based Sybil defenses are based on either random walks or community detections. Random Walk Based Sybil Classification: Sybil Guard and Sybil Limit were the first schemes to propose Sybil detection using social network structure. Sybil Limit relies on the insight that social networks are relatively well connected, and thus short random walks starting from benign users can quickly reach all other benign users. The intersection of random walks is used as a feature by the benign users to validate each other. On the other hand, short random walks from Sybil users do not reach all benign users (due to limited number of attack edges), and thus do not intersect with the random walks from benign users. Random Walk Based Sybil Ranking: Sybil Rank performs random walks starting from a set of benign users. Specifically, with h labeled benign nodes, Sybil Rank designs a special initial probability distribution over the nodes, i.e., probability 1/h for each of the labeled benign nodes and probability 0 for all other nodes, and iterates the random walk from this initial distribution for log(n) iterations, where n is the total number of nodes in the network.CIA is also based on a random walk with a special initial probability distribution, but it differs from Sybil Rank in two major aspects. First, CIA starts the random walk from labeled malicious nodes. Second, in each iteration, CIA restarts the random walk from the special initial probability distribution with some probability.

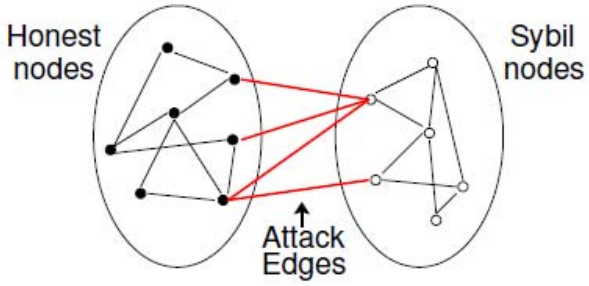


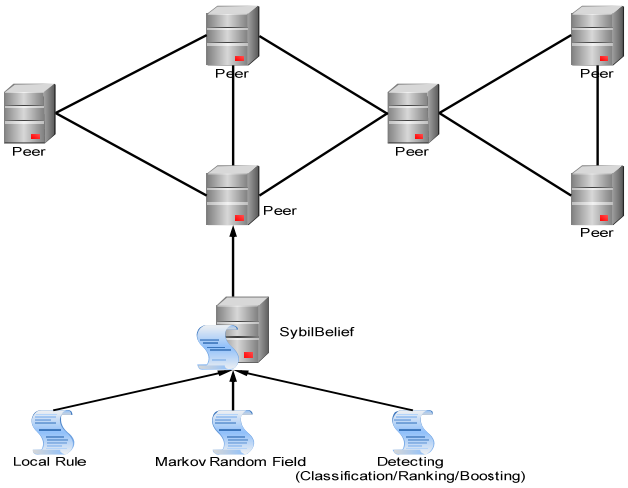
Figure 1: The social network with honest nodes and sybil nodes. Note that regardless of which nodes in the social network are sybil nodes, we can always "pull" these nodes to the right side to form the logical network in the figure.

III. Problem Defination

A semi-supervised learning framework, to detect Sybil nodes. SybilBelief takes asocial network of the nodes in the system, a small set of known benign nodes, and, optionally, a small set of known Sybils as input. Then, SybilBelief propagates the label information from the known benign and/or Sybil nodes to the remaining nodes in the system. We evaluate SybilBelief using both synthetic and real-world social network topologies. We show that SybilBelief able to accurately identify Sybil nodes with low false positive rates and low false negative rates. SybilBelief is resilient to noise in our prior knowledge about known benign and Sybil nodes. a semi-supervised learning problem, where the goal is to propagate reputations from a small setoff known benign and/or Sybil users to other users along the social connections between them. More specifically, we first associate a binary random variable with each user in the system; such random variable represents the label (i.e., benign or Sybil) of the user. Second, we model the social network between users in the system as a pair wise Markov Random Field, which defines a joint probability distribution for these binary random variables. Third, given a set of known benign and/or Sybil users, we infer the posterior probability of a user being benign, which is treated as the reputation of the user.

An undirected social network $G = (V, E)$, where a node $v \in V$ represents a user in the system and an edge $(u, v) \in E$ indicates that the users $u \in V$ and $v \in V$ are socially connected. In an ideal setting, G represents a weighted network of trust relationships between users, where the edge weights represent the levels of trust between users. Each node is either *benign* or *Sybil*. We denote the subnet work including the benign nodes and the edges between them as the *benign region*, denote the subnet work including the Sybils and edges between them as the *Sybil region*, and denote the edges between the two regions as *attack edges*. two linked nodes have the same label. Online social network operators can obtain social networks that satisfy homophily via two methods. One method is to approximate *trust* relationships between users through looking into user interactions inferring tie strengths and asking users to rate their social contacts.

The other method is to preprocess the networks so that they are suitable for structure based approaches. In particular, operators could first detect and remove compromised benign nodes (e.g., front peers), which decreases the number of attack edges and increases the homophily. Showed that some simple detectors might enforce the social networks to be suitable for structure-based Sybil defenses if the attack edges are established randomly.



IV. SYBILBELIEF MODEL

To introduce our approach SybilBelief, which is scalable, tolerant to label noise, and able to incorporate both known benign labels and Sybil labels.

4.1 An Example:

To quantify the homophily in social networks, we first propose a new probabilistic local rule which determines the reputation score for a node v via aggregating its neighbors' label information. Then, we demonstrate that this local rule can be captured by modeling social networks as Markov Random Fields (MRFs). Specifically, each node in the network is associated with a binary random variable whose state could either be *benign* or *Sybil*, and MRFs define a joint probability distribution over all such random variables. Given a set of known benign labels and/or known Sybil labels, the posterior probabilities that nodes are benign are used to classify or rank them. We adopt Loopy Belief Propagation to approximate the posterior probabilities

4.2 Design Goals

Our goal is to detect Sybils in a system via taking a social network between the nodes in the system, a small set of known benign nodes, and (optionally) a small set of known Sybils as input. Specifically, we have the following design goals.

1) Sybil Classification/Ranking: Our goal is to design a mechanism that can either classify nodes into benign and Sybil or that can rank all nodes in descending order of being benign.

2) Incorporating Known Labels: In many settings, we already know that *some* users are benign and that some users are Sybil. For instance, in Twitter, verified users can be treated as known benign labels and users spreading spam or malware can be

treated as known Sybil labels. To improve overall accuracy of the system, the mechanism should have the ability to incorporate information about both known benign and known Sybil labels. It is important that the mechanism should not *require* information about known Sybil labels, but if such information is available, then it should have the ability to use it. This is because in some scenarios, for example when none of the Sybils have performed an attack yet, we might not have known information about any Sybil node.

3) Tolerating Label Noise: While incorporating information about known benign or known Sybil labels, it is important that the mechanism is resilient to noise in our prior knowledge about these labels. For example, an adversary could compromise the account of a known benign user, or could get a Sybil user whitelisted. We target a mechanism that is resilient when a minority fraction of known labels are incorrect.

4) Scalability: Many distributed systems (e.g., online social networks, reputation systems) have hundreds of millions of users and billions of edges. Thus, for real world applicability, the computational complexity of the mechanism should be low, and the mechanism should also be parallelizable.

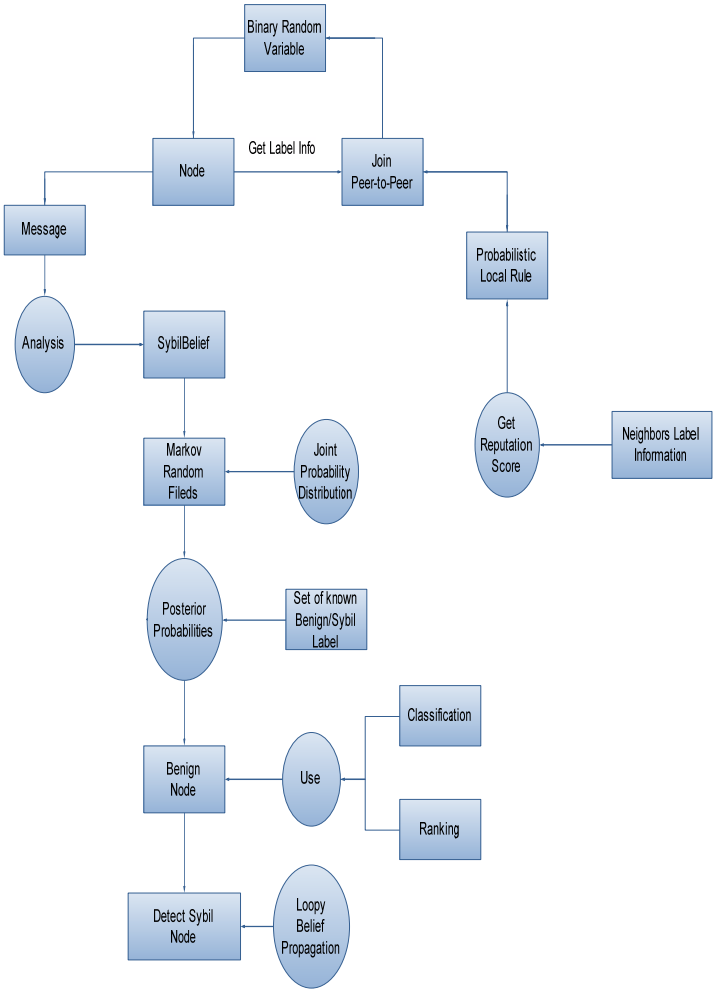


Fig.3 Overall Dataflow Diagram

V. SYBILBELIEF LEARNING ALGORITHM

Our Sybil classification and ranking mechanisms rely on the computation of the posterior distributions given in Equation 6. Generally, there are two major ways to infer such posterior distributions: *sampling* and *variational inference*. We adopt variational inference to learn the posterior distributions since it is more scalable than sampling approaches such as Gibbs sampling. Specifically, we adopt Loopy Belief Propagation (LBP) to calculate the posterior distributions for each node.

5.1 Loopy Belief Propagation (LBP):

The basic step in LBP is to pass messages between neighboring nodes in the system. Message *m(t) uv (xv)* sent from *u* to *v* in the *t*th iteration is

m(t) uv (xv) =_xuφLu (xu)ϕuv (xu, xv) _k∈_(u)/v m(t−1) ku (xu)

Here, *_(u)/v* is the set of all neighbors of *u*, except the receiver node *v*. This encodes that each node forwards a product over incoming messages of the last iteration and adapts this message to the respective receiver based on the coupling strength with the receiver.

For social networks without loops (i.e., for trees), LBP is guaranteed to converge and to compute the exact posterior distribution. For networks with loops, LBP approximates the posterior probability distribution without convergence guarantees. However, in practical applications and benchmarks in the machine learning literature, LBP has demonstrated good results and is, today, a widely used technique.

5.2 Stopping Condition: The message passing iterations stop when the changes of messages become negligible (e.g., L1 distance of changes becomes smaller than 10−3) or the number of iterations exceeds a predefined threshold. After stopping, we estimate the posterior probability distribution

P(xv | . xL) by *P(xv | . xL) ∝ Φl v (xv) _ k∈_(v) m(t) kv (xv)*

5.3 Scalability*:* The complexity of one LBP iteration is *O(m)*, where *m* is the number of edges. So the total complexity is *O(m ∗ d)*, where *d* is the number of LBP iterations. Note that social networks are often sparse graphs. Thus we have *O(m ∗ d) = O(n ∗ d)*, where *n* is the number of nodes. Moreover, we find that setting *d* to be 10 already achieves good results in our experiments. Furthermore, LBP can be easily parallelized. Specifically, we can distribute nodes in the system to multiple processors or computer nodes, each of which collects messages for nodes assigned to them.

5.4 Pairwise Markov Random Field

We find that the probabilistic local rule introduced in the previous section can be applied by modeling the social network as a pairwise Markov Random Field (MRF). A MRF defines a joint probability distribution for binary random variables associated with all the nodes in the network. Specifically, a MRF is specified with a *node potential* for each node *v*, which incorporates prior knowledge about *v*, and with an *edge potential* for each edge *(u, v)*, which represents correlations between *u* and *v*. In the context of Sybil detection, we define a node potential *φv (xv)* for the node *v* as

φv (xv) :=_θv if *xv* = 11 − *θv* if *xv* = −1

and an *edge potential ϕuv (xu, xv)* for the edge *(u, v)* as

ϕuv (xu, xv) := _wuv if *xuxv* = 11 − *wuv* if *xuxv* = −1,

where *θv := (*1+exp{−*hv* }*)*−1 and *wuv := (*1 +exp{−*Juv* }*)*−1.

Then, the following MRF satisfies the probabilistic local rule.

P(xV) = 1 *Z_v∈Vφv (xv) _(u,v)∈Eϕuv (xu, xv),*

where *Z = _xV _v∈V φv (xv)_(u,v)∈E ϕuv (xu, xv)* is called the partition function and normalizes the probabilities

Summary*:* We have the following observations:

» SybilBelief accepts more Sybil nodes in the PA-generated networks than in the ER-generated networks. This implies that attackers should design their Sybil regions to approach scale-free networks.

» SybilBelief is robust to label sites.

» There exists a phase transition point $w0$ (e.g., $w0 \approx 0.65$ in our experiments) for the parameter w. SybilBelief performance is robust for $w > w0$.

» SybilBelief only requires one label per community.

» SybilBelief can tolerate 49% of labels to be incorrect. Moreover, SybilBelief can detect incorrect labels with 100% accuracy.

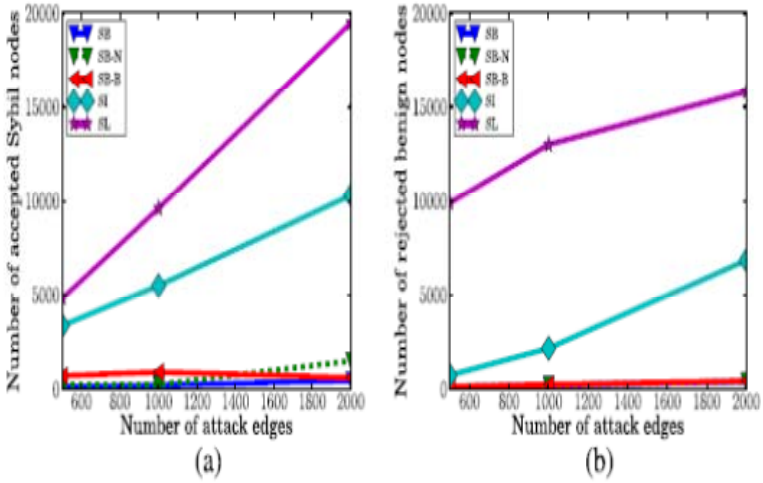
» SybilBelief is robust to community structures in the benign region.

» With only benign or Sybil labels, our boosting strategy can still achieve performances comparable to the case where both benign and Sybil labels are observed. Furthermore, the number of boosting trials can be used to balance between accepted Sybil nodes and rejected benign nodes.

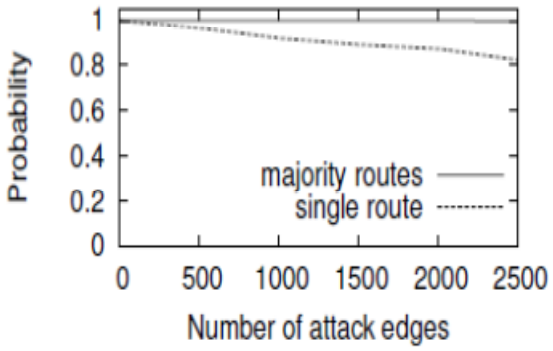
TABLE-1

NOTATIONS OF ALGORITHMS

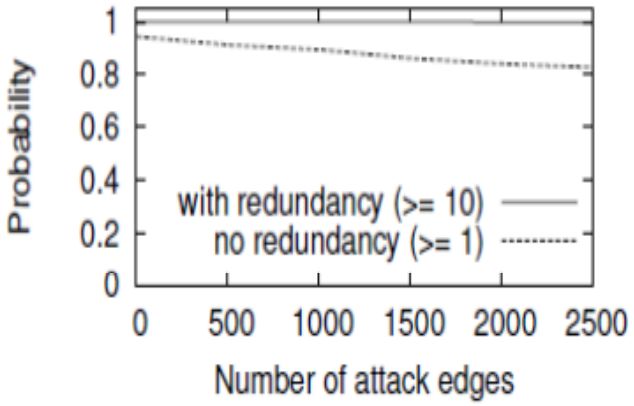
Notation	Description
SL	SybilLimit [9]
SI	SybilInfer [10]
SR	SybilRank [13]
SR-N	SybilRank [13] with label noise
CIA	Criminal account Inference Algorithm [14]
CIA-N	CIA with label noise
SB	SybilBelief
SB-N	SybilBelief with label noise
SB-B	SybilBelief with only labeled benign nodes
Random	Randomly assign a reputation score between 0 and 1 to each node



The number of accepted Sybil nodes and rejected benign nodes as a function of the number of attack edges. The benign region is the Facebook network, and the Sybil regions are synthesized by PA model. We observe that SB, SB-N, and SB-B all work an order of magnitude better than previous classification systems. Furthermore, we find that incorporating both benign and Sybil labels increases the performance of our algorithm. (a) Accepted Sybil nodes. (b) Rejected benign nodes.



Probability of routes remaining entirely within the honest region.



Probability of an honest node accepting another honest node (i.e., having at least a targenumber of intersections).

VI. Conclusion

In SybilBelief, a semi-supervised learning framework, to detect Sybil nodes in distributed systems.SybilBelief takes social networks among the nodes in the system, a small set of known benign nodes, and, optionally, a small set of known Sybil nodes as input, and then SybilBelief propagates the label information from the known benign and/or Sybil nodes to the remaining ones in the system. We extensively evaluate the influence of various factors including parameter settings in the SybilBelief, the number of labels, and label noises on the performance of SybilBelief. Moreover, we compare SybilBelief with state-of-the-art Sybil classification and ranking approaches on real-world social network topologies. Our results demonstrate that SybilBelief , performs orders of magnitude better than previous Sybil classification mechanisms and significantly better than previous Sybil ranking mechanisms. Furthermore, SybilBelief is more resilient to noise in our prior knowledge about known benign nodes and known Sybils.

References

1. J. R. Douceur, "The Sybil attack," in Proc. 1st Int. Workshop Peer-to- Peer Syst., 2002.
2. K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, "Design and evaluation of a real-time URL spam filtering service," in Proc. IEEE Symp. Security Privacy, May 2011, pp. 447–462.
3. L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda, "All your contacts are belong to us: Automated identity theft attacks on social networks," in Proc. 18th Int. Conf. WWW, 2009, pp. 551–560.
4. B. Viswanath, A. Post, K. P. Gummadi, and A. Mislove, "An analysis of social network-based Sybil defenses," in Proc. SIGCOMM, 2010,pp. 363–374.

Author(s) Profile

Dr. K. Swathi obtained her under-graduation in B.E., (Computer Science & Engineering) from Bharathidasan University, Trichy in 1999. She obtained her M.E. degree in Computer and Communication Engineering from Anna University, Chennai in 2004.She obtained Ph.D degree in Faculty of Information & Communication Engineering from Anna University, Chennai in 2014. Presently, she is working as Associate Professor in Computer Science & Engineering, Cauvery College of Engineering and Technology, Trichy in 2008. She has 14 years teaching experience and also she had attended many workshops, seminars and conferences on Research issues in Image processing. She has published papers in international journals and presented papers in various Conferences. She is the life member of Indian Society for Technical Education (ISTE). Her areas of interest include Image processing, Data mining, network security and Software engineering.

P.Nidhya was born in Thanjavur, Tamilnadu, India, in 1989. She received the B.Tech. degree in Information technology from Anjalai ammal mahalingam engineering college, Tiruvarur affiliated to Anna University Trichirapalli, India, in 2011, and Currently she is pursuing her M.E. degree in computer science and engineering at cauveri Engineering college, Trichy affiliated to Anna university, Chennai., India, Her area of interest is information security.

Vijayanathan.R. received his Master of philosophy Library and Information Science from Annamalai University in 1999 Also he obtained his post-graduate degree in Master of Economics in Bharathidasan University in 1996 and PG Diploma in Computer Application, PG Diploma Tourism Management and PG Diploma in co- operative management from Annamalai Unversity .He is working as a Sr. Librarian, Department of library and Information Science in Cauvery College of Engineering and technology, Trichy from 2009. He has guided many M.Phil scholars produced and member in various Universities in Tamil Nadu. He served more than 14 years as Senior Librarian in reputed Engineering College and also he had attended many workshops, seminars and conferences on Research issues in Bibliometric ananalysis.He has published so many research papers in National and International journals. His areas of interested are networking; Cloud computing, IC Technologies, Environmental study, Library Automation, Webometric study, Scientometric study, Bibliometric analysis, and citation study.

