Compromised Account Detection using Friends Closiveness in Online Social Networks

Vaslit.J[1], S. Arulkumar[2]

[1]III MCA, Lord Jegannath College of Engineering & Technology

[2]Assistant Professor, Department of Computer Applications,

Lord Jegannath College of Engineering & Technology

Abstract:- Compromised accounts in Online Social Networks (OSNs) are more positive than Sybil accounts to spammers and other malicious OSN attacker. Malicious parties exploit the well-established relations and trust connection between the genuine account owners and their friends. In this paper we study the friend's closiveness feature of OSN users, i.e. their usage of OSN services, and the application of which in detecting compromised accounts. We propose a set of social friend's closiveness features that can effectively characterize the user social activities on OSNs. We validate the efficacy of these friends' closiveness features by collecting and analyzing real user click streams to an OSN website. Based on our measurement study, we devise individual user's social behavioral profile by combining its respective friends colsiveness features metrics. A social behavioral profile correctly reflects a user's OSN activity patterns. While an authentic owner conforms to its account's social behavioral profile reluctantly, it is hard and costly for imposters to feign.

I. INTRODUCTION

Compromised accounts in Online Social Networks are more positive than Sybil accounts to spammers and other malicious OSN attackers. Malicious parties develop the well-established connections and trust relationships between the legitimate account owners and their friends, and efficiently distribute spam ads, phishing links, or malware, while avoid being blocked by the service providers. Offline analyses of tweets and Facebook posts [1], [2] reveal that most spam are distributed via compromised accounts, instead of dedicated spam accounts. Recent large-scale account hacking incidents [3], [4] in popular OSNs further evidence this trend.

Unlike dedicated spam or sybil accounts, which are created solely to serve malicious purposes, compromised accounts are originally possessed by benign users, While dedicated malicious accounts can be simply banned or removed upon detection, compromised accounts cannot be handled likewise due to potential negative impact to normal user experience (e.g., those accounts may still be actively used by their legitimate benign owners). Major OSNs today employ IP geo location logging to battle against account compromization. [5], [6].However, this approach is known to suffer from low detection granularity and high false positive rate. Previous research on spamming account detection [7], [1],[2], [8] mostly cannot distinguish compromised accounts from Sybil accounts, with only one recent study by Egele*et al.* [9] features compromised accounts detection. Existing approaches involve account profile analysis [10], [8], and message content analysis [7], [9], [2], [11], [12] (e.g. embedded URL analysis [2], [11] and message clustering [9],[9]). However, account profile analysis is hardly applicable for detecting compromised accounts, because their profiles are the original common users' information which is likely to remain intact by spammers. URL blacklisting has the challenge of timely protection and update, and message clustering introduces significant overhead when subjected to a large number of real-time messages. Instead of analyzing user profile contents or message contents, we seek to uncover the behavioral anomaly of compromised accounts by using their legitimate owners' history social activity patterns, which can be observed in a lightweight manner. To better serve users' various social communication needs, OSNs provide a great variety of online features for their users to engage in, such as building connections, sending messages, uploading photos, browsing friends' latest updates, etc. However, how a user involves in each activity is completely driven by personal interests and social habits. As a result, the interaction patterns with a number of OSN activities tend to be divergent across a large set of users. While a user tends to conform to its social patterns, a hacker of the user account who knows little about the user's behavior habit is likely to diverge from the patterns. Therefore, as long as an authentic user's social patterns are recorded, checking the compliance of the account's upcoming behaviors with the authentic patterns can detect account compromization. Even though user's credential is hacked; a malicious party cannot easily obtain the user's social behavior patterns without the control of the physical machines or the click streams. Moreover, considering that for a spammer, who carries very different social interests from those of regular users (e.g., mass spam distribution vs. entertaining with friends), it is very costly to mimic different individual user's social interaction patterns, as it will significantly reduce spamming efficiency.

II. RELATED WORKS

Schneider *et al.* [13] and Benevento *et al.* [14] measured OSN users' behaviors based on network traffic collected from ISPs. Both works analyze the popularity of OSN services, session length distributions, and user click sequences among OSN services, and discover that browsing accounts for a majority of users' activities. Benevento *et al.* [14] further explored user interactions with friends and other users multiple hops away. While these works primarily emphasize the

overall user OSN service usage, and aim to uncover general knowledge on how OSNs are used, this paper studies users' social behavior characteristics for a very different purpose. We investigate the characterization of individual user's social behaviors to detect account usage anomaly. Moreover, we propose several new user behavioral features and perform measurement study ata fine granularity. Viswanath*et al.* [15] also aim to detect abnormal user behaviors in Facebook, but they sole focus on "like" behaviors to detect spammers. While most previous research on malicious account detection cannot differentiate compromised accounts from spam accounts, Egele*et al.* [9] specifically studied the detection of compromised accounts.

By recording a user's message posting features, such as timing, topics and correlation with friends, they detected irregular posting behaviors; on the other hand, all messages in certain duration are clustered based on the content, and the clusters in which most messages are posted by irregular behaviors are classified as from compromised accounts. While they also leveraged certain user behavior features to discern abnormality, we use a different and more complete set of metrics to characterize users' general online social behaviors, instead of solely focusing on message posting behaviors.

Additionally, our technique does not rely on deep inspection and classification of message contents and avoids the heavy weight processing. Wang *et al.* [16] proposed an approach for Sybil account detection by analyzing click streams. They differentiated Sybil and common users' clicks based on inter-arrival time and click sequence, and found that considering both factors leads to better detection results. Since Sybil's are specialized fake identities owned by attackers, their click stream patterns significantly differ from those of normal users. However, for compromised accounts, their click streams can be a mix from normal users and spammers, As a result, methods in [16] cannot handle compromised accounts well.

In contrast, this paper aims to uncover users' social behavior patterns and habits from the click streams, with which we can perform accurate and delicate detection on behavioral deviation. Regarding spammer detection, [17] and [18] set up honey pot accounts to harvest spam and identify common features among spammers, such as URL ratio in their messages and friends choice; using those features, both employ classification algorithms to detect spammers. Yang *et al.* [8] introduced new features of spammers involving with their connection characteristics to achieve better accuracy. Thomas *et al.* [11] analyzed the features of fraudulent accounts bought from the underground market and developed a classifier using the features to retrospectively detect fraudulent accounts. Instead of focusing on malicious accounts, Xie*et al.* [19] proposed to vouch normal users based on the connections and interactions among legitimate users. As for spam detection, Gao*et al.* [7] proposed a real-time spam detection system, which consists of a cluster recognition system to cluster messages and a spam classifier using six spam message features. Thomas *et al.* [11] thrived to detect spam by identifying malicious URLs in message content. In [1], [2], the authors conducted offline analysis to characterize social spam in Facebook and Twitter, respectively. They found that a significant portion of spam was from compromised accounts, instead of spam accounts. Meanwhile, Yang *et al.* [20] investigated connections among identified spammers and other malicious account detection methods [21], [22], [23] exploit the differences on static profile or connectivity information between normal and malicious accounts.

III. COMPROMISED ACCOUNT DETECTION FRAMEWORK

We proposed several new friends closiveness features that can effectively quantity user differences in online social activity. For each behavioral feature, we deduce a behavioral metric by obtaining a statistical distribution of the value ranges, observed from each user's clickstream. Moreover we combine the respective behavioral metrics of each user into a social behavioral profile, which represents a user's social behavioral

pattern. To validate the effectiveness of social behavioral profile in detecting account activity anomaly, we apply the social behavioral profile of each user to differentiate click stream of its respective user from all other users. We conduct multiple cross-validation experiments , each with varying amount of input data for building social behavioral profile can effectively differentiate individual OSN user with none accuracy. Previous works focused on all the social behaviors, but our proposed friends closiveness features will immediate higher user behaviors based on best close friends. This friend's closiveness feature set improves the accuracy of the compromised account detection.

1. FRIENDS CLOSIVENESS DETECTION

We classify user social behaviors on an OSN into two classes, extroversive behaviors and internal behaviors. External behaviors, such as uploading photos and sending messages, result in able to be seen imprints to one or more peer users; internal behaviors, such as browsing other users' profiles and inquiring in message inbox, however, do not create observable effects to other users. While most earlier research only focus on the external behaviors, such as public posting [8], we study both classes of behaviors for a more complete understanding and description of user social behaviors.

2. FRIENDS CLOSIVENESS BEHAVIOYRAL PROFILE

The first external activity a user engages in after logging in an OSN session can be habitual. Some users regularly start from commenting on friends' new updates; while some others are more inclined to update their own rank first. The first activity feature aims to capture a user's ordinary action at the starting of each OSN session. How often a user engages in each type of extroversive activities relates to their personalities [5]. Some users like to post photos, while some others spend more time response to friends' posts; some mostly chat with friends via private messages, while some others always speak by posting on each other's public message boards. Typical OSNs offer a great variety of social actions to gratify their users' communication needs, for example, commenting, updating status, posting notes, sending messages, sharing posts, inviting others to an event, etc. As a result, this feature can give a detailed portrayal of a user's social communication preferences. The relative order a user completes multiple external activities.

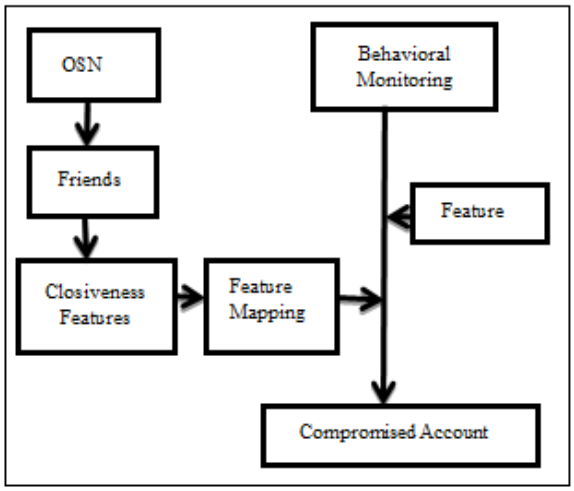


Fig 1: Architecture Diagram

While users have their preferences on different social activities, they may also have habitual patterns when switch from one action to another. For instance, after commenting on friends' updates some users often update their own status, while some other users prefer to send messages to or chat with friends instead. Therefore, the action series feature reflects a different social behavioral pattern from the activity preference. The speed of actions when a user engages in certain external activities reflects the user's social interaction style.

Many activities on OSNs require multiple steps to finish. For example, posting photos involves loading the upload page, selecting one or more photos, uploading, editing (e.g., clipping, decorating, tagging, etc.), previewing and verification. The time a user takes to complete each action of a given activity is seriously influenced by the user's social characteristics (e.g.,

serious vs. casual) and knowledge with the respective activity; but it doesn't directly reflect how fast a user acts due to different content complexity. The *action latency* feature is proposed to provide more fine-grained and accurate metric. Although invisible to peer users, introversive behaviors make up the majority ofa user's OSN activity; as studied in previous work [6], [15]the dominant (i.e., over 90%) user behavior on an OSN is browsing. Through introversive activities users gather and consume social information, which helps them to form ideas and opinions, and eventually, establish social connection sand initiate future social communications. Hence, introversive behavior patterns make up an essential part of a user's online social behavioral characteristics. We propose the following four features to portray a user's introversive behavior.

3. FRIENDS CLOSIVENESS FEATURES MAPING

We first conduct a systematic study of services and web pages on Facebook. Based on request URL, we categorize 29 different types of extroversive activities that a user can conduct to interact with peer users; we also classify 9 types of Facebook web pages containing different kinds of social information, which users can browse privately(i.e., the introversive activities). With the mapping between the click stream information and the user behaviors, we analyze each user's click streams to extract the corresponding behavior patterns. We present the combined measurement results of each behavior feature for all users to show the value space, and finally we use an example to illustrate user behavior diversities.

4. COMPROMISED ACCOUNT DETECTION

The social behavioral profile depicts various aspects of a user's online social behavior patterns, and it enables us to quantitatively describe the differences in distinct user social behaviors. In the following, we first describe how to compare social behavioral profiles by calculating their difference. Then, we argue the application of social behavioral profile relationship to individual different users and detecting compromised accounts.

1) Comparing Behavior Profiles:

Given any two social behavioral profiles, P and Q, we quantify their *difference* in two steps. In the first step, we compare each of the eight vectors in Paganism the respective vector in Q. Particularly; we measure the Euclidean distance to quantify the difference between the two vectors. Given two vectors A = (a1, a2, ...an) and B = (b1, b2, ...bn), the Euclidean distance between them is calculated by,

$$E(A,B) = \sqrt{\sum_{t=1}^{n} (ai - bi) 2}$$

Comparing all eight vectors yield an eight-element Euclidean distance vector (E1,E2, ...,E8). Each element in this vector has a range of [0,√2], because the sum of each vector's elements is one. In the second step, we take the Euclidean norm of the Euclidean distance vector,

$$D(P,Q) = \sqrt{\sum_{j=1}^{8} (Ej) 2}$$

The resulting value is the difference of the two behavioral profiles, and has a range of [0, 4]—the more significant the two profiles differ, the larger the value is.

IV. ANALYSIS

We first verify that behavioral profile can accurately portrayal user's behavior pattern. Next, we validate the feasibility of employing behavioral profiles to distinguish different users, which can be used to detect compromised accounts.

A. Detection Accuracy

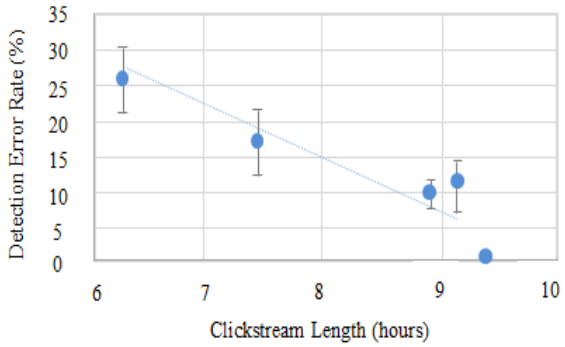
Here we further evaluate the accuracy of using social behavioral profiles to differentiate online users. We conduct three sets of experiments by varying training data size, feature quality, and profile completeness, respectively, to evaluate their Calculated from its click stream, and other users are taken as impostors. For each impostor, we calculate its

behavioral profile difference to U's using U's weights on each feature. If the difference is larger than VU + 2 ∗stdDev(U), the decision is taken as correct; otherwise, it is taken as an error. The average error rate is calculated in each scenario. Setting the threshold to beVU + 2 ∗stdDev(U) guarantees that U's behavioral profile scan be discerned with the probability of more than 97%.

1. **Input Size vs. Accuracy:**

Intuitively, the more training data are given to build a user's behavioral profile, the better the profile reflects its behavior pattern; hence the profile difference demonstrates the dissimilarity between two user behaviors more accurately.

We build each user's behavioral profile using the click stream from 1/6... and 1/2 of its total sessions, respectively, and use cross-validation to compute and compare of behavioral profile differences.



(a) Training Data Size vs. Accuracy

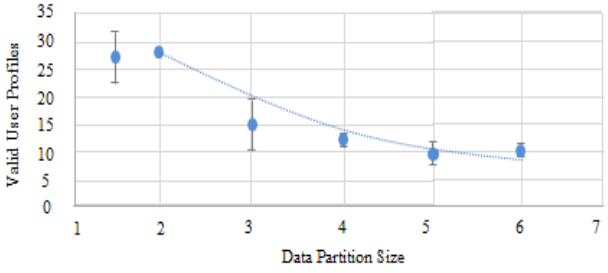


Fig . 2: Impact of Training Data Size

(b) Valid Users w. Training Data Size

Take the 1/6 of sessions as an example, each user's click stream is partitioned into 6 parts while the first part includes the click stream from the 1st, 7th, 13th, ...,sessions; the second part includes the click stream from the2nd, 8th, 14th, ..., sessions etc. Six behavioral profiles are built accordingly and each profile is used for difference calculation. For user A, when we use each part of its click stream to build its behavioral profile, the behavioral profile difference from another user B to user A is calculated six times, each of which considers A's behavioral profile and one of B's behavioral profiles, which is built from one out of its 6 click streams. Cross-validation is used to make sure that each part of data is used for both training and validation, and the result is not derived from biased data. Furthermore, we only consider users whose behavioral profiles consist of more than or equal to 4 non-empty feature vectors, each of which should be built from more than or equal to 15 sample activities. The thresholds are set to guarantee the vector quality as well as the completeness of

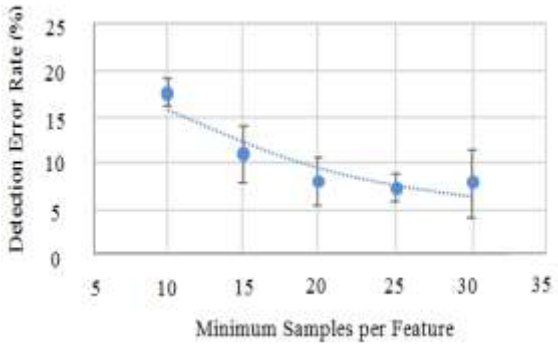
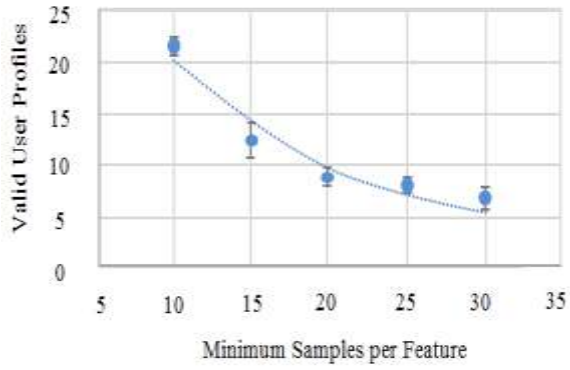


Fig. 3: Impact of Feature Quality

(a) Feature Quality vs. Accuracy



(b) Valid Users w. Feature Quality the behavioral

profiles.

To give more straight forward impression over the size of click stream we calculate its average active hours in each partition. Figure 2(a) shows the dynamics of the error rate with the change of the click stream length, while Figure 2(b) shows the number of valid users with different data partitions. Overall, the longer is the click stream, the more accurate is the detection. When the click stream is up to 8.9 hours, the error rate can reach 9.5% while longer click streams derive better result. When it is more than 9.3 hours, the error rate is as small as0.3%. Longer click stream provides more empirical behavior data of a user, which enables us to build more accurate and complete behavioral profiles, and hence the distance on each behavioral feature can be measured more accurately. Thus, the detection is more accurate. On the other hand, with the finer partition of the click stream, the fewer activities in each partition, leading to the smaller number of non-empty vectors and the smaller number of valid users. The side-effect to set a high sample threshold is that with limited click stream, fewer feature vectors can be built, resulting in fewer valid users, as Figure 3(b) shows.

2. Profile Completeness vs. Accuracy:

Due to the lack of certain activities, some behavioral feature vectors can be N/A. For instance, when one never conducts extroversive activities in its click stream, at least four of its feature vectors are N/A, which makes its profile incomplete. By adjusting the least number of non-empty features vectors, the completeness of selected behavioral profiles can be guaranteed.

V. CONCLUSION

Our evaluation on sample Facebook users indicate that we can achieve high detection accuracy when behavioral profiles are built in a complete and accurate fashion. We presented a novel approach to detect compromised accounts in social networks. More precisely, we developed statistical models to characterize the behavior of social network users, and in the future we used anomaly detection techniques to identify sudden changes in their behavior.

REFERENCES

1. H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao. Detecting and characterizing social spam campaigns. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, IMC '10, pages 35–47, Melbourne, Australia, 2010. ACM
2. K.-I. Goh and A.-L. Barab´asi. Burstiness and memory in complex systems. *EPL (Europhysics Letters)*, 81(4):48002, 2008.
3. 250,000 twitter accounts hacked. http://www.cnn.com/2013/02/01/tech/social-media/twitter-hacked.
4. 50,000 facebook accounts hacked. http://www.ktsm.com/news/ thousands-of-facebook-accounts-hacked.
5. Detecting suspicious account activity. http://googleonlinesecurity.blogspot.com/2010/03/detecting-suspicious-account-activity.html.
6. Facebook tracks the location of logins for better security. http:// www.zdnet.com/blog/weblife/facebook-adds-better-security- tracks- the location- of-your-logins/2010.
7. H. Gao, Y. Chen, and K. Lee. Towards online spam filtering in social networks. In *Symposium on NetowrkandDistributed System Security,* NDSS 12', San Diego, CA USA. Internet Society.
8. C. Yang, R. C. Harkreader, and G. Gu. Die free or live hard? Empirical evaluation and new design for fighting evolving twitter spammers. In *Proceedings of the 14th international conference on Recent Advances in Intrusion Detection*, RAID'11, pages 318–337, Menlo Park, CA, 2011. Springer-Verlag.
9. M. Egele, G. Stringhini, C. Kruegel, and G. Vigna. Compa: Detecting compromised accounts on social networks. In *Symposium on Network andDistributed System Security*, NDSS 13', San Diego, CA USA. Internet Society.
10. K. Thomas, D. McCoy, C. Grier, A. Kolcz, and V. Paxon. Trafcking fraudulent accounts: The role of the underground market in twitter spam and abuse. In *Proceddings of 22nd USENIX Security Symposium*, USENIX Security 13', Washington D.C., USA, 2013.

11. K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song. Design and evaluation of a real-time url spam filtering service. In *IEEE Symposium on Security and Privacy*, S&P 11', pages 447–462, Oakland, CA, USA,2011. IEEE Computer Society.
12. D. Wang, D. Irani, and C. Pu. Evolutionary study of web spam: Webb spam corpus 2011 versus webb spam corpus 2006. In IEEE International Conference on Collaborative Computing: Networking, Applications and*Worksharing*, CollaborateCom '12, pages 40–49. IEEE, 2012.
13. F. Schneider, A. Feldmann, B. Krishnamurthy, and W. Willinger. Understanding online social network usage from a network perspective. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, IMC '09, pages 35–48, Chicago, Illinois, USA,2009. ACM.
14. F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida. Characterizinguser behavior in online social networks. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, IMC '09, pages 49–62, Chicago, Illinois, USA, 2009. ACM.
15. B. Viswanath, M. A. Bashir, M. Crovella, S. Guha, K. P. Gummadi, Krishnamurthy, andA. Mislove. Towards detecting anomaloususer behavior in online social networks. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 223–238, San Diego, CA, Aug. 2014.USENIX Association.
16. G. Wang, T. Konolige, C. Wilson, X. Wang, H. Zheng, and B. Y. Zhao.You are how you click: Clickstream analysis for sybil detection. In *Proceddings of 22nd USENIXSecurity Symposium*, USENIX Security 13', Washington D.C., USA, 2013.
17. G. Stringhini, C. Kruegel, and G. Vigna. Detecting spammers on social networks. In *Proceedings of the 26th Annual Computer SecurityApplications Conference*, ACSAC '10, pages 1–9, Austin, Texas, USA, 2010. ACM.
18. K. Lee, J. Caverlee, and S. Webb. Uncovering social spammers: socialhoneypots + machine learning. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in informationretrieval*, SIGIR '10, pages 435–442, Geneva, Switzerland, 2010. ACM.
19. Y. Xie, F. Yu, Q. Ke, M. Abadi, E. Gillum, K. Vitaldevaria, J. Walter, J. Huang, and Z. M. Mao. Innocent by association: early recognition of legitimate users. In *ACM Conference on Computer and CommunicationsSecurity*, CCS '12, pages 353–364, Raleigh, NC, USA, 2012. ACM.
20. C. Yang, R. Harkreader, J. Zhang, S. Shin, and G. Gu. Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter. In *Proceedings of the 21st internationalconference on World Wide Web*, WWW '12, pages 71–80, Lyon, France, 2012.ACM.
21. Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro. Aiding the detection of fake accounts in large scale social online services. In *Proceedings of the 9th USENIX conference on Networked Systems Design and Imple-mentation*, NSDI'12, San Jose, CA, USA, 2012. USENIX Association.
22. J. Song, S. Lee, and J. Kim. Spam filtering in twitter using senderreceiver relationship. In *Proceedings of the 14th international conference on Recent Advances in Intrusion Detection*, RAID'11, pages 301–317,Menlo Park, CA, USA, 2011. Springer-Verlag.
23. B. Viswanath, A. Post, K. P. Gummadi, and A. Mislove. An analysis of social network-based sybil defenses. In *Proceedings of the ACM SIGCOMM 2010 conference*, SIGCOMM '10, pages 363–374, NewDelhi, India, 2010. ACM.