Spam Source Clustering by Constructing Spammer Network with Correlation Measure

Jeongkyu Shin and Seunghwan Kim

Asia Pacific Center for Theoretical Physics
Nonlinear and Complex System Laboratory,
Pohang University of Science and Technology, Pohang 790-784, Korea
jkshin@physics.postech.ac.kr, swan@postech.ac.kr

Abstract. Spam filtering is one of the most challenging problems in electric message systems. In general, recent studies on specifying real spam source are based on content filtering because spammers usually falsify their origin. We propose a method to specify spam source based on structural analysis with complex network. We assume that each spam sources either has the same victim list or uses the same spam-hosting program. We treat spam source - target relationship as a bipartite network and construct weighted spam source network by network projection using correlation measure. We find that community clustering methods are inappropriate with spammer network. We group spammers with gradient-based grouping, which uses correlations between nodes as gradient between nodes. We convert them into local minima, which helps to cluster spammers into a few spam source groups. We investigate the weblog spam data with the proposed method and validate it. The method that we propose can be applied to diverse categorization problems, such as multiple text categorization and network subunit clustering.

Keywords: Electronic spam, complex network, clustering method.

Undesired electronic messages on World-Wide Web (for instance, E-Mail, comments, trackbacks, and etc.) are becoming a serious problem of electric communication. These 'spam' messages are usually sent for advertising services and products, and some of them contain malwares that is made for cracking receivers' computers.

Spam and spam-blocker development is a good example of fast evolving system with competition. Spam usually does not contain helpful information for receivers. At the beginning of online advertisement market, spam could be treated as noise in electronic communication because the portion of spam was relatively small. Now the situation changed: most electronic messages are spams. Spam consumes traffic, which is strongly related with communication resource. As many studies are performed to avoid spam, spams also start to evolve: it is a stiff fight between a spear and a shield.

Blacklist is one of well-known traditional filtering methods. If a message is marked as a spam, spam filter adds the signature (usually URL or IP address)

J. Zhou (Ed.): Complex 2009, Part I, LNICST 4, pp. 889–893, 2009.

to blacklist. Blacklist method is easiest and strongest filtering method. However IP spoofing disturbs finding the spam sources [1]. Spam senders (called *Spammers*) usually deceive spam referrer IP using zombie computers which are hacked by spammers, thus the number of spammer looks quite big even though it is small [2].

A modern approach, *Bayesian spam filtering*[3], based on Bayes theorem has become a common filtering method due to its flexibility and effectiveness. Bayesian filtering is one of *content pattern matching methods* which requires more computing resource than blacklist. In spite of the power of Bayesian filtering, Spammers use some tricks to evade spam filtering. For instance, *Bayesian poisoning* is a well-known method to spoil Bayesian filtering. Basically, modern spam filtering consists of combining blacklist and Bayesian filtering which are common spam filtering methods.

IP spoofing exaggerates the number of spammers, and IP blacklist can be too complex as a side effect. By the reason that spammer has content or URL pattern for their spam, grouping zombie computers and specifying spammer can help content-based filtering filter to extract spam pattern. Thus specifying exact spam sources helps developing anti-spam methods.

In this paper, we propose a method to group diverse spammer IPs based on complex network analysis with *spam source - target relation* data. Specifying spam source is usually based on content (including URL, IP address) analogy. We suggest a hypothesis that even though spammers use zombie computers to spread spam messages, same spammer uses the same target list. If targets are almost the same, those attacks can be originated from the same spam source (or same spamming program). If this hypothesis is true, spammers might be grouped by reconstructing spammer-spammer relation network even if spammers falsify their IP address. We treat *spam source - target relationships* as bipartite network [4], and project spam-target vector into spam sources relation matrix using correlation measures.

Bipartite network structure can be converted into unipartite one [5]. Let $\bar{\sigma}^{\mu}$ as signature vector of spammer IP address μ, where

$$\bar{\sigma}_i^{\mu} = \begin{cases} 1 \text{ if spam attacks victim } i \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

Note that every spammer and victims are labeled. $\bar{\sigma}^{\mu}$ denotes the record which victims are attacked by spammer μ. Thus the information of victim network layer is treated as the connection information of spammer network. After construction of every spammer's signature vector, relations between spammer IP addresses can be calculated with similarity measure which is a kind of symmetric correlation measure [6]

$$C^{\mu\lambda} = \frac{\bar{\sigma}^{\mu} \cdot \bar{\sigma}^{\lambda}}{|\bar{\sigma}^{\mu}||\bar{\sigma}^{\lambda}|} \equiv \cos\theta_{\mu\lambda} \tag{2}$$

where $\bar{\sigma}^{\mu} \cdot \bar{\sigma}^{\lambda}$ is the scalar product between signature vectors. We can calculate every correlation between all pairs of spammers and construct spammer-spammer weighted adjacency network C.

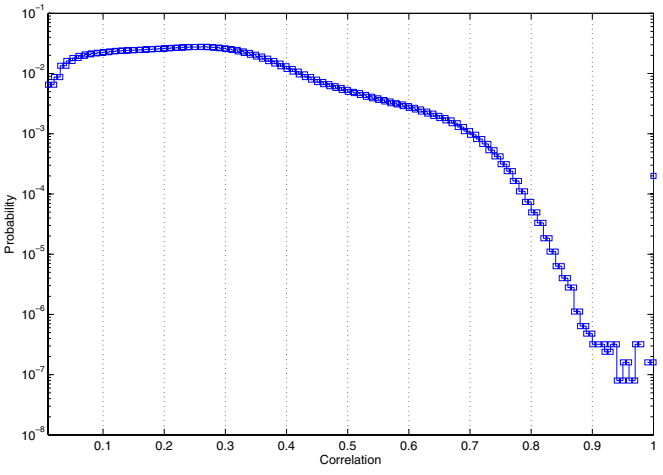

Fig. 1. Probability distribution of spammer network link weight. Result indicates that the percolation-based filtering does not work well with weblog spammer network.

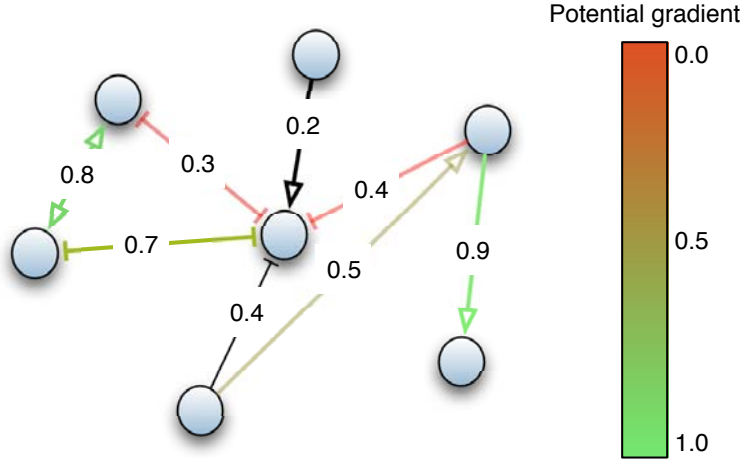

Fig. 2. Gradient-based grouping method example. Each node makes group with another node with highest link weight. Note that this graph is not an directed graph; arrow indicates the belongings, not link direction.

Once the weighted network is determined, various group methods are studied. Percolation-based filtering[4], which is the commonly used grouping method by disconnecting nodes lower than specific filtering coefficient value, has a problem with weblog spam network. Fig.1 shows that there is no transition point as correlation varies. Thus it is hard to determine the proper filtering coefficient.

We propose *gradient-based grouping method*, inspired from energy potential landscape network[7]. Gradient-based grouping method works as follows: first,

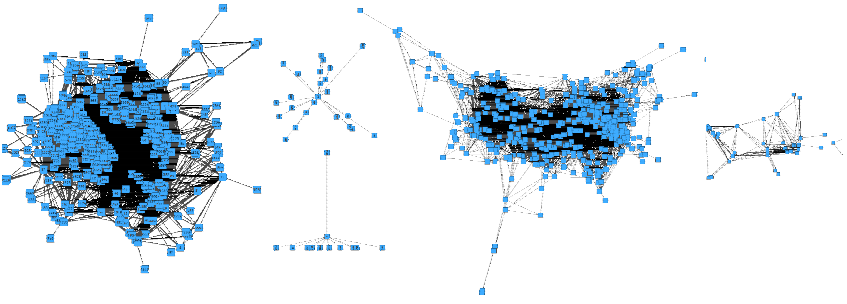

Fig. 3. Percolation-based filtering result. Filtering coefficients are *left*) 0.68 and *right*) 0.75. Groups are made by breaking biggest island.

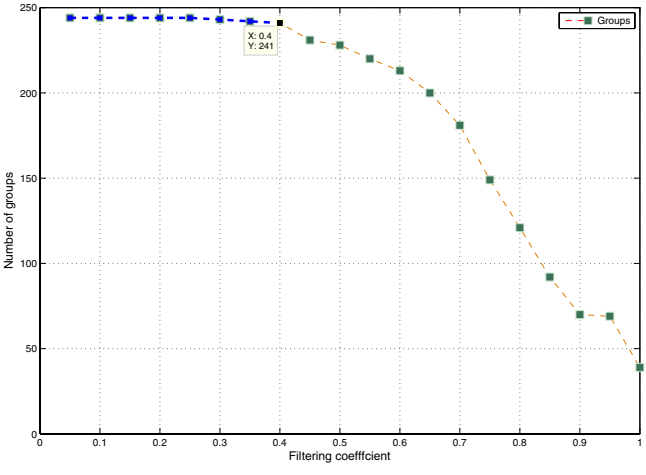

Fig. 4. Relation between number of groups and filtering coefficient using gradient-based group method. Result shows many groups and various group sizes. However, there is no dominant group unlike other methods. The number of groups is not affected by filtering coefficient.

check the weight of links of node i. Find link and connected node j with maximum weight. Finally, mark node i as a same group with j. Repeat these process for every nodes. Time complexity is $\sim O(n)$ for finding link with maximum weight, and $O(n)$ for loop. Thus the time complexity is around $\sim O(n^2)$.

We investigate the weblog (commonly called as *blog*) spam data from Eolin anti-spam service (EAS) [8]. We use spam data between July. 2006 and Nov. 2006. We extract 3118667 unique trackback spam pairs from spammer-victim IP and prune 500th to 5500th spammers (861328 pairs) to reduce side-effect from data irregularity. We find that the number of weblog spammers is usually bigger

than number of victims similar to the E-mail spam, where spammers are known to counterfeit the sender information.

We extract *spam source IP - spam target* relationship as a vector sets, and construct the spam source adjacency matrix. Percolation-based grouping shows the linear fragmentations as filtering coefficient varies. Groups are made by breaking biggest island, thus the number of island heavily depends on the filtering coefficient (Fig. 3). In contrast, result from gradient-based grouping method shows uniformly sized, same number of group even if the filtering coefficients varies (See Fig. 4).

To validate our grouping method, we compare the result with those from other common methods. We perform content-based grouping by extracting representative words from spam content. By comparing the content-based grouping with the gradient-structure-based grouping, result shows that 81% of groups are same. We emphasize that, even content-based grouping and our network structure based IP grouping are totally different methods, test results show meaningful similarities; this result supports out hypothesis about spam target lists.

In conclusion, we address a grouping method in spam source specifying problem. We propose a methodology to group spam sources, which helps analyzing spam patterns and characterizing the properties of them, based on complex network analysis. We construct spammer IP address network from a spam source-target bipartite network using correlation measures, and classify spammer groups with gradient-based grouping method. We validate our method by comparing with other methods.

Our method can be applied to diverse categorization problems with sparse data, such as multiple text categorization, time-series analyses and network subunit clustering.

Acknowledgments. Raw data and some tests for this study has been supported by *Tatter and Company (TNC)*, acquired by Google inc. at Sep. 2008.

References

1. Spamlinks.net, `http://spamlinks.net/filter-bl.htm`
2. Song, S., Manikopoulos, C.N.: IP Spoofing Detection Approach(ISDA) for Network Intrusion Detection System. In: Sarnoff Symposium. IEEE, Los Alamitos (2006)
3. Sahami, M., Dumais, S., Heckerman, D., Horvitz, E.: A Bayesian Approach to Filtering Junk E-Mail. In: AAAI 1998 Workshop on Learning for Text Categorization (1998)
4. Newman, M.E.J., Strogatz, S.H., Watts, D.J.: Random graphs with arbitrary degree distributions and their applications. Phys. Rev. E. 64, 026118 (2001)
5. Newman, M.E.J.: The structure of scientific collaboration networks. Proc. Natl. Acad. Sci. U.S.A. 98, 404–409 (2001)
6. Lambiotte, R., Ausloos, M.: Uncovering collective listening habits and music genres in bipartite networks. Phys. Rev. E. 72, 066107 (2005)
7. Doye, J.P.K.: The network topology of a potential energy landscape: A static scale-free network. Phys. Rev. Lett. 88, 238701 (2002)
8. Eolin Antispam Service, `http://antispam.eolin.com`