# Revealing Social Networks of Spammers

Kevin S. Xu
Alfred O. Hero III

Spam doesn't really need an introduction—anyone who owns an email address likely receives spam emails every day. However, spam is much more than just an annoyance. Spam's hidden economic cost for companies in wasted storage, bandwidth, technical support, and most important, the loss of employee productivity, is astronomical. The annual cost of spam for a company with 12,000 employees is approximately $2.4 million, according to a study conducted by *Windows & .NET Magazine* in 2003 [1]. Since then, the amount of spam received has only increased. According to estimates from MessageLabs, over 80 percent of emails received from 2005 to 2008 were spam [2].


New
F7
6
7
8
T
Y
U

The magnitude of the spam problem has not gone unnoticed by the US government. In 2003, the United States government drafted the Controlling the Assault of Non-Solicited Pornography and Marketing (CAN-SPAM) Act to address the issue. CAN-SPAM provided guidelines on unsolicited email practices and specified how unsolicited email could be sent legally. Unfortunately, compliance has been extremely low; therefore, the act has had virtually no effect on lowering the volume of spam.

On the other hand, CAN-SPAM allowed Internet service providers (ISPs) and web site owners to file lawsuits against spammers, resulting in fines and occasional jail sentences for convicted spammers. While lawsuits are certainly a way to fight back against spammers, given the vast number of spammers, suing an individual has a negligible effect on reducing the overall volume of spam, especially when lawsuits are brought regardless of the impact of the offense. Unfortunately, spammers have responded by taking greater measures to conceal their identities to avoid being detected. Clearly, other mechanisms are necessary to combat spam effectively.

One type of spam that represents a significant threat to individuals and companies alike is *phishing* spam. Phishing is an attempt to fraudulently acquire sensitive information by appearing to represent a trustworthy entity. Phishing spam often takes the form of emails appearing to be from a trusted financial institution with which the recipient does business. These emails are written to persuade the recipient to reveal confidential information such as online banking passwords, credit card numbers, or a social security number. Many victims of identity theft have been fooled into revealing sensitive information by phishing emails.

Current methods to combat spam before it reaches a user include *content-based filtering* at the recipient's email server as well as *blacklisting* email servers known to send only spam emails. Both measures reduce the annoyance of spam and the loss of employee productivity by decreasing spam emails arriving at employee inboxes. However, these strategies can also backfire. For example, content-based filtering has the unintended side effect of misclassifying legitimate email as spam.

Furthermore, filtering does nothing to reduce the volume of spam that is sent. When spammers know that a smaller percentage of emails are getting past the spam filters to the intended recipients, they might compensate by sending more spam emails. Thus, content-based filtering may even increase the volume of spam sent!

Email servers that send only spam can be blacklisted to filter out all emails sent from them. Blacklisting differs from content-based filtering in that the filtering is done on email servers instead of on individual emails. Blacklisting is a more efficient filtering approach, but the disadvantage to blacklisting is that many email servers send both legitimate email as well as spam; blacklisting such a server would result in legitimate emails being misclassified as spam.

Current anti-spam methods share one common weakness—they are local; that is, they detect and filter out spam at a single location, which is the recipient's email server. Local anti-spam solutions are easy to maintain because a single administrator, usually the information technology group of the company or ISP, manages the process. But what could an analyst discern by examining how spam operates on a greater network level?

In this article, we investigate the spam problem using a global approach, which requires detection and monitoring of an entire network or at multiple locations within a network. By taking a global approach, an analyst can correlate data over multiple email servers, times, and locations to infer the behavior of spammers on a large scale, which can then be used to combat spam nearer to its source.

The best defense spammers have against anti-spam techniques is to send spam emails without being detected. So how do they do this? Consider the path of spam, illustrated in Figure 1. First, a spammer acquires email addresses on a web page using a *harvester*, which is a piece of software designed to visit web sites and extract email addresses from the HTML source code. Next, spam servers send emails to the acquired addresses. These can be servers that belong to the spammers, or they can be *zombie* computers, computers compromised by viruses or other malware that end up sending spam without their owners' knowledge. Finally, these spam emails make their way to the recipients' inbox or junk mail folder.

The address acquisition process, known as *harvesting*, is an often overlooked part of the spam problem. Malicious spammers typically take

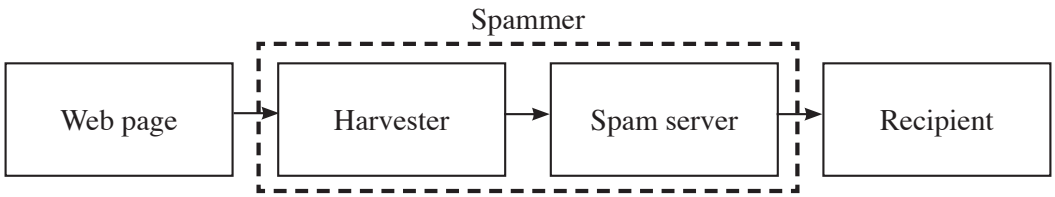


Figure 1: The path of spam from an email address on a web page to your inbox

measures to conceal their identities when sending spam. One common method is to use massive networks of compromised computers, known as *botnets*. However, studies have indicated that spammers do not take comparable precautions when harvesting [3], perhaps because harvesting is seen as a safer and more acceptable activity than sending spam. Hence, monitoring harvesting activity and tracking harvesters can be useful for identifying spammers. This is one of the goals of Project Honey Pot, created by anti-spam company Unspam Technologies, Inc. [4].

## Project Honey Pot

Project Honey Pot was started in 2004 to monitor harvesting and spamming activity via a network of decoy web pages set with trap email addresses, known as *honey pots*. These honey pots are embedded in the HTML source code of a web page and are invisible to human visitors. Harvesters looking for email addresses in HTML source code sometimes stumble across the trap addresses and acquire them. Harvesters can also be directed to trap addresses by links to honey pots from legitimate web sites that they also scan for email addresses.

Each time a honey pot is visited, the centralized Project Honey Pot server generates a unique trap email address. The visitor's IP address is associated with the trap email address and then recorded on the server. The email address embedded in the honey pot is unique, so only the visitor to that honey pot could have collected it. Because these trap email addresses are not published anywhere besides the honey pot, all emails received at these addresses are assumed to be spam.

Project Honey Pot provides a unique opportunity to investigate the social structure of spammers. It is normally very difficult to uncover anything at the spammer level because we cannot associate a spam email with a particular spammer. The "from" address can be easily spoofed, and the spam served from a compromised computer has little association with the spammer. With Project Honey Pot each spam email is associated with the harvester that acquired the recipient's email address. When spammers fail to conceal their identities while harvesting, the IP address of the harvester is likely to be closely related to the actual location of the spammers.

Because each email received at a trap email address is associated with the harvester that acquired it, the identity of the spammer is revealed. As of March 2010, Project Honey Pot comprised over 48 million honey pots distributed all over the world [4]. The data collected by Project Honey Pot provides a global perspective on spam and makes it possible to investigate correlations over many spam servers and time periods.

## Discovering communities of spammers

As mentioned earlier, understanding the behavior of spammers on an expanded scale is one of the benefits of a global approach for fighting spam. But what do the social networks of spammers look like? In particular, how well organized are spammers? Do they operate alone, or in groups? Are there meaningful communities or organizations of spammers? Sending spam emails is profitable for spammers; otherwise, there wouldn't be so much spam. Can a business model be derived from the community structure of spammers? These questions can be answered using the data collected by Project Honey Pot and a technique known as *spectral clustering* [5].

The social network of spammers can be represented as a graph consisting of nodes and edges, as shown in Figure 2. The nodes correspond to spammers, and an edge between two nodes corresponds to a social relationship between the corresponding spammers. A social relationship can be inferred by the use of common resources or by similar behavior patterns over time. Communities in a social network emerge by partitioning the graph into groups of nodes. Sets of nodes in the same group are highly similar and sets of nodes in different groups are not similar. Spectral clustering

aims to minimize the normalized cut between groups, which is defined by

$$\text{Normalized cut} = \frac{\text{Sum of all edge weights between groups}}{\text{Sum of all edge weights within groups}}$$

For example, spectral clustering divides the graph shown in Figure 2 into the two communities indicated by the blue and green nodes, respectively. The groups revealed by spectral clustering correspond to communities in the social network. For these communities to be meaningful, the graph must be constructed so the edges between nodes correspond to actual relationships between spammers.

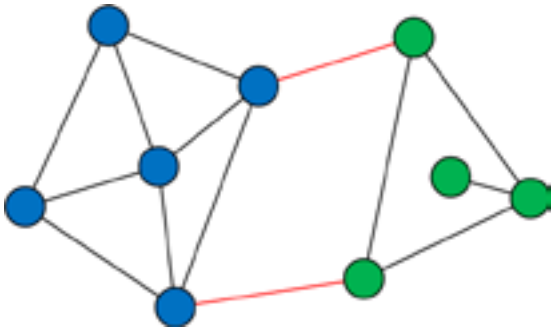

Figure 2: An example of a graph and its separation into two communities by spectral clustering

The main challenge in constructing the graph is choosing the edges and edge weights, because we cannot observe relationships among spammers. This problem does not arise in most other community detection studies. For example, in friendship or collaboration networks, users willingly participate in the study, and information on relationships among members of the network is readily available. However, for spammer network discovery, relationships between spammers are only inferred through correlations between behavior patterns. Two spammers who have high behavioral correlation are likely to be collaborating. This likelihood, which is treated as the strength of the relationship between these two spammers, can be used as the weight of the edge between the two corresponding nodes in the graph. For this research, we investigate two types of behavioral correlation between spammers: correlation in spam server usage and temporal correlation.

## Correlation in spam server usage

Correlation in spam server usage between two spammers corresponds to common usage of a set of spam servers. Spammers typically try to conceal their identity by using spam servers that aren't traceable back to them, such as botnets. Thus spam servers can be viewed as resources for spammers, and common usage of a set of spam servers between two spammers translates into resource sharing, which suggests that the two spammers are collaborating. By constructing the graph using correlation in spam server usage between all active spammers over a period of time, many interesting communities of spammers are revealed, as shown in Figure 3.

Each node in the graph corresponds to a spammer, and the color and shape of a node indicates the community to which he or she belongs. Note that the majority of spammers belong in a large, loosely-connected community identified by the red nodes. These are the spammers who do not exhibit extremely high correlation with other spammers. Hence it is not a true community, but a collection of spammers who appear to be operating alone. The interesting communities are the smaller, tightly-connected ones surrounding the large red community. We believe that these nodes correspond to actual social communities of spammers working together and sharing substantial email server resources.

Reinforcing our belief is the observation that the discovered communities tend to divide into phishing and non-phishing communities, as shown in Figure 4. The shade of each node corresponds to the phishing level of each spammer, which is defined by

$$\text{Phishing level} = \frac{\text{Number of phishing emails sent}}{\text{Total number of emails sent}}$$

We denote spammers with high phishing levels as *phishers* and the rest as *non-phishers*. Notice that phishers tend to form communities with other phishers, and that non-phishers tend to form communities with other non-phishers. This is also evident from looking at the most frequent subject lines of emails from all spammers in a community. For example, the most frequent subject lines from both a phishing community, namely the orange community of triangular nodes at the top of Figure

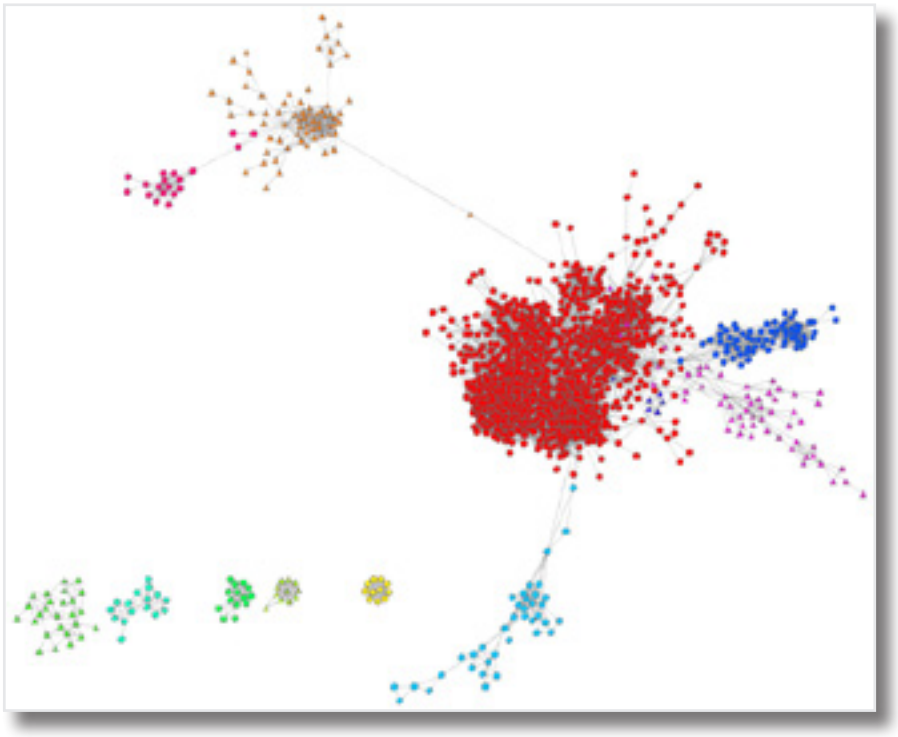

Figure 3: Community structure of spammers inferred by correlation in spam server usage in October 2006

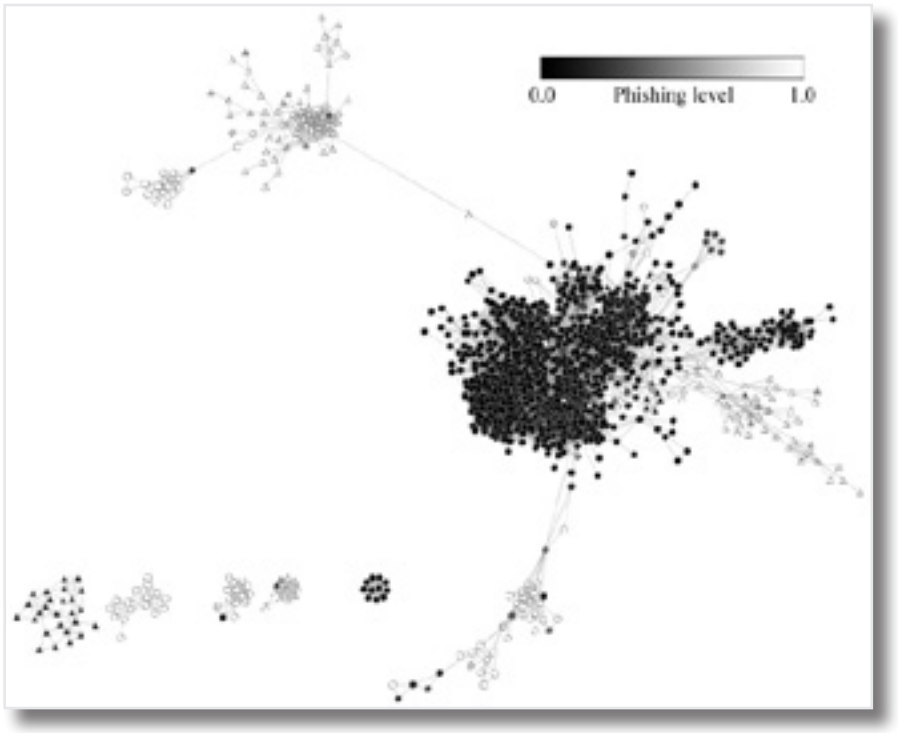


Figure 4: Alternate view of the same social network shown in Figure 3, shaded by phishing level

| Phishing Community | Non-Phishing Community |
|---|---|
| Password Change Required | Make Money by Sharing Your Life with Friends and F |
| Question from eBay Member | Premiere Professional & Executive Registries Invit |
| Credit Union OnlineÂ® $50 Reward Survey | Texas Land/Golf is the Buzz |
| PayPal Account | Keys to Stock Market Success |
| PayPal Account - Suspicious Activity | An Entire Case of Fine Wine plus Exclusive Gift to |

Table 1: Most common subject lines from a phishing and a non-phishing community (truncated to 50 characters by the Project Honey Pot database)

3, and a non-phishing community, namely the blue community of circular nodes on the right of Figure 3, are listed in Table 1. Notice the distinct separation between phishing subject lines and non-phishing subject lines. The subject headings were not provided to the clustering algorithm and therefore confirm that server usage patterns alone can provide evidence of coordinated phishing behavior. We note that phishers tend to concentrate in small, tightly-connected communities. This observation provides empirical evidence that communities of phishing spammers are sharing resources, namely spam servers, among the community. This suggests that phishers tend to exist in isolated, well-organized social communities or teams.

## Temporal correlation

*Temporal correlation* refers to correlation of the times when emails were sent. High correlation is expected among spammers who are working together. Because we do not know the times when emails were sent, we correlate the times when emails were received. The community structure as revealed using temporal correlation is shown in Figure 5.

Again, the shape and color of a node represent the community that a particular spammer belongs to. Two large communities appear, and as before, they can be interpreted as loosely-connected communities of individuals who do not exhibit much correlation with each other. However, in the smaller communities, some interesting patterns emerge. In particular, we discover groups of spammers with nearly coherent temporal spamming behavior. Consider the group of ten spammers whose temporal spamming behavior is shown in Figure 6, in which the horizontal axis corresponds to days in a month and the vertical axis corresponds to the number of emails sent each day. The figure

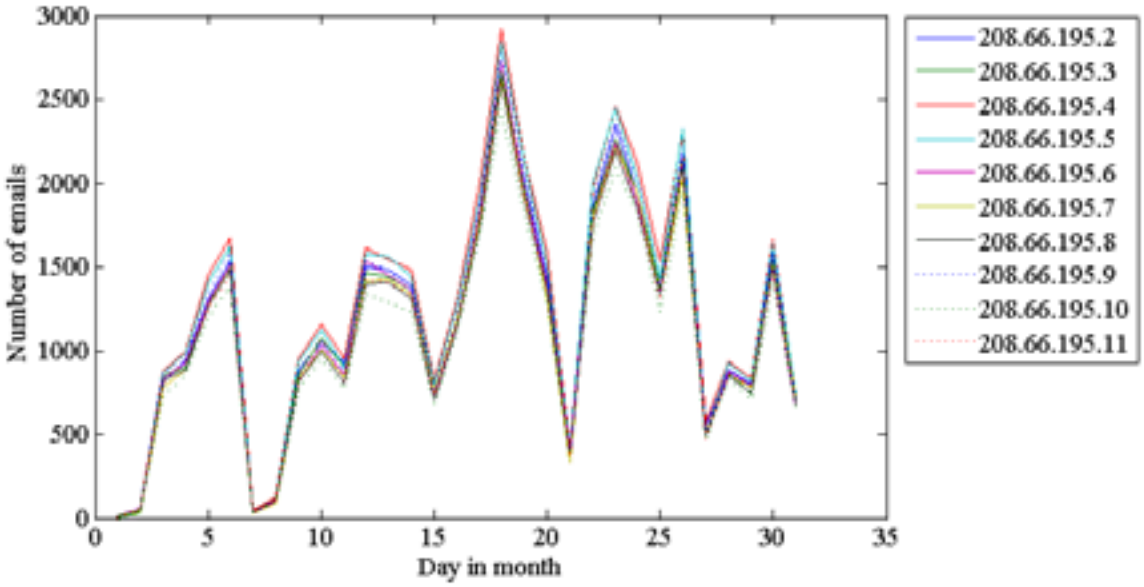


Figure 6: Temporal spamming behavior of group of ten spammers over the month of October 2006, by IP

consists of ten lines overlaid onto the same plot, with each line corresponding to the temporal spamming behavior of one spammer in the group.

How striking that the ten spammers in Figure 6 are sending almost identical numbers of emails over time! And how probable that they are working together and belong to an actual social community. These ten spammers, found in the community of dark-blue colored nodes in the top left of Figure 5, are especially interesting because they are among the heaviest spammers in the Project Honey Pot data set, where a heavy spammer denotes someone who sends a large number of spam emails. In addition to their highly coherent temporal behavior, these spammers also have IP addresses in the same block, indicating that they are operating from the same physical location, perhaps in the same building. Furthermore, these ten spammers' IP addresses are in the IP address range of a known rogue ISP, McColo Corp., which had been hosting and providing services for cybercriminals until it was taken down in November 2008 [6]. All of the abovementioned observations point to this group of spammers being very well-organized, and thus we conclude that they form a tight social community.

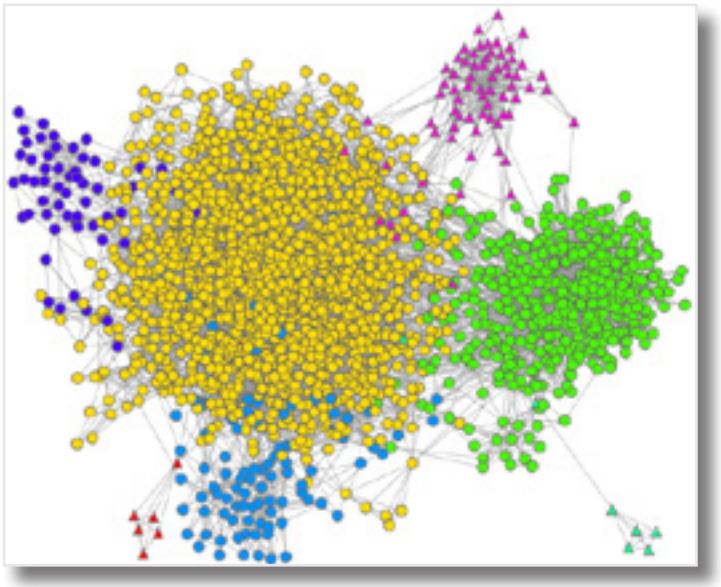

Figure 5: Community structure of spammers inferred by temporal correlation in October 2006

## Conclusions

Current methods of fighting spam are local and take place at the receiving end, which does not help to reduce the amount of network traffic consumed by spam emails. By studying spam from a global perspective using the data collected by Project Honey Pot, we were able to correlate the behavior of spammers, allowing us to identify different communities of spammers. We found that the majority of spammers appeared to be working alone, but a significant number of them appear to form communities or organizations. In particular, we discovered many small communities of spammers who predominantly sent phishing emails, likely attempting to acquire sensitive information to engage in identity theft. We also discovered several communities of spammers operating from the same physical location, suggesting strong social connections between these spammers.

By analyzing spam and spammer behavior from a global perspective, we were able to identify meaningful communities of spammers. The next step would be to use these findings to combat spam. Several avenues that could be pursued include identifying social cliques that could perhaps be linked to an organization and identifying important members of the social network who could be sued, which would have a much greater effect than randomly targeting spammers. There is also potential for online detection of communities; that is, updating the detected communities as emails are received. This would allow for a new method of spam filtering, not by content or blacklisting, but by behavioral patterns of spammers, which are less variable. Thus filtering by behavioral patterns has the potential to be more effective than existing filtering methods.

Although the problem of spam does not appear to be going away anytime soon, methods and tools for combating it are improving. Spectral clustering and network discovery can lead to insights into how spammers operate by revealing their social networks. The methods described in this paper might also be applied to discovery of illicit behavior patterns in other applications, such as financial transaction networks or chat room interaction networks. For additional details on our methods, the reader is referred to "Revealing Social Networks of Spammers Through Spectral Clustering" [7].

## Acknowledgment

The authors would like to thank Dr. Mark Kliger, Yilun Chen, and Dr. Peter Woolf for their contributions to the presented analysis. We also thank Matthew Prince, Eric Langheinrich, and Lee Holloway of Unspam Technologies for providing us with the Project Honey Pot data set. This research was partially supported by Office of Naval Research grant N00014-08-1-1065 and National Science Foundation grants CCF 0830490 and CCR-0325571.

## About the Authors

**Kevin S. Xu** is a PhD student in the Department of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor, working under the supervision of Dr. Hero. His current areas of interest are in the study of networks, including information networks and social networks among others. In particular, he is interested in developing methods for inferring properties of networks such as their community structure or their temporal dynamics from observations taken within the network.

**Alfred O. Hero III** is the R. Jamison and Betty Williams Professor of Engineering at the University of Michigan, Ann Arbor, where he is affiliated with the Department of Electrical Engineering and Computer Science, the Department of Biomedical Engineering, and the Department of Statistics. His recent research interests have been in signal extraction and pattern analysis for Internet, sensor net, and bionet data streams. He is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE) and has received an IEEE Signal Processing Society Meritorious Service Award (1998), best paper awards from the Clinical Cytometry Society (2009) and the IEEE Signal Processing Society (1998, 2010), and the IEEE Third Millennium Medal (2000). Alfred Hero was President of the IEEE Signal Processing Society (2006-2008) and sits on the Board of Directors of the IEEE as Director of Division IX (2010-2011).

## References:

[1] The Secret Cost of Spam. *Windows & .NET Magazine*. [Online]. Available: http://www.itmanagement.com/whitepaper/the-secret-cost-of-spam/, 2003.

[2] MessageLabs Intelligence: 2008 Annual Security Report. Symantec Corp. [Online] Available: http://www.messagelabs.com/download.get?filename=MLIReport_Annual_2008_FINAL.pdf.

[3] M. Prince, L. Holloway, E. Langheinrich, B. M. Dahl, and A. M. Keller. Understanding How Spammers Steal Your E-Mail Address: An Analysis of the First Six Months of Data from Project Honey Pot. In *Proc. 2nd Conf. Email and Anti-Spam*, 2005.

[4] Project Honey Pot. Unspam Technologies Inc. [Online] Available: http://www.projecthoneypot.org/, 2009.

[5] S. Yu and J. Shi. Multiclass Spectral Clustering. In *Proc. 9th IEEE Int. Conf. Computer Vision*, 2003.

[6] J. Nazario. Third 'Bad ISP' Disappears—McColo Gone. Arbor Networks. [Online] Available: http://asert.arbornetworks.com/2008/11/third-bad-isp-dissolves-mccolo-gone/, 2008.

[7] K. S. Xu, M. Kliger, Y. Chen, P. J. Woolf, and A. O. Hero III. Revealing Social Networks of Spammers Through Spectral Clustering. In *Proc. IEEE Int. Conf. Communications*, 2009.