



## City Research Online

### City, University of London Institutional Repository

---

**Citation:** Abu-Nimeh, S., Chen, T. & Alzubi, O. (2011). Malicious and spam posts in online social networks. *Computer*, 44(9), pp. 23-28. doi: 10.1109/mc.2011.222

This is the published version of the paper.

This version of the publication may differ from the final published version.

---

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/8204/>

**Link to published version:** <https://doi.org/10.1109/mc.2011.222>

**Copyright:** City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

**Reuse:** Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.



# Malicious and Spam Posts in Online Social Networks

Saeed Abu-Nimeh, *Damballa Inc.*

Thomas M. Chen and Omar Alzubi, *Swansea University, Wales*

**A large-scale study of more than half a million Facebook posts suggests that members of online social networks can expect a significant chance of encountering spam posts and a much lower but not negligible chance of coming across malicious links.**

**T**he popularity of online social networks has been growing exponentially. Launched in February 2004, Facebook—the world’s largest social network—had 250 million active users by July 2009 but doubled that number within just one year. According to Facebook’s own statistics ([www.facebook.com/press/info.php?statistics](http://www.facebook.com/press/info.php?statistics)), the average user has 130 friends and creates 90 pieces of content—news stories, blog posts, notes, photos, hyperlinks, and so on—monthly. The total user population spends 700 billion minutes on Facebook and shares more than 30 billion pieces of content each month.

The most obvious threat to users in social networks is loss of privacy. In July 2010, a security researcher revealed that the account details of more than 100 million Facebook accounts were publicly accessible through search engines.<sup>1</sup> In addition to loss of privacy, social network users face spam and various malicious threats including social engineering, identity theft, browser exploits, and malware.

Hackers target online social networks for several reasons:

- such networks contain large target populations;
- there is an abundance of personal information to steal or exploit;

- joining is fairly easy;
- users tend to have a high level of trust in one another and for objects (messages, links, photos, applications) within the networks;
- the variety of shared Web content, including hyperlinks and applications, exposes users to a range of potential attack vectors; and
- social graphs are highly interconnected, offering the potential for viral dissemination of malware and other attacks—a property of human social networks made famous by the “six degrees of separation” postulated by social psychologist Stanley Milgram.<sup>2</sup>

Despite the popularity and widely recognized security risks of online social networks, there have been very few large-scale investigations of the real extent of malicious threats. The “Related Work” sidebar summarizes the general findings of these efforts.

To assess the prevalence of malicious and spam posts in Facebook, we analyzed more than half a million posts with the help of Defensio, a Facebook application that protects users from such content as well as filters profanity and blocks URL categories. Our analysis revealed that a significant fraction of Facebook posts is spam and a much smaller fraction is malicious.

## FACEBOOK ARCHITECTURE

Typical for online social networks, Facebook is designed to allow a community of users to easily share information, messages, links, photos, and videos. After filling in a profile page, users can choose various levels of information access for different visitors. In addition, users can establish connections to designate “friends” or join groups. They can also send messages to one another through the Social

## RELATED WORK

**M**ost work on the risks of online social networks has focused on privacy concerns, but numerous researchers have looked at security threats.

Lynn Greiner noted that many attacks exploit the implicit trust between users in a social network, which makes people more likely to click on fake links or fall for social engineering schemes.<sup>1</sup>

Weimin Luo and colleagues surveyed numerous general threats to social networks and identified various attacker motivations and attack vectors—namely, spam, applications, malware, Web vulnerabilities, browser plug-ins, and social engineering.<sup>2</sup>

A team led by Tom Jagatic showed that it is easy to use data from social networks to hone phishing attacks.<sup>3</sup> They discovered that the success rate of phishing increases dramatically when e-mail appears to come from friends. Supporting this point, Garrett Brown and colleagues<sup>4</sup> found that, although Facebook itself does not reveal users' e-mail addresses, most such addresses can be obtained through public databases linked to the network. Furthermore, on most publicly accessible Facebook profiles, contextual information is available that hackers could exploit to generate context-aware spam. These researchers discovered that even a fraction of users with closed (private) profiles is vulnerable to such spam.

Several studies have considered the security risks related to Facebook applications.

Andrew Besmer and coauthors pointed out that Facebook app users are initially asked for permission to allow access to their profile data, but even if they do not consent, an app can still request such data on behalf of a friend who installed it.<sup>5</sup>

Constantinos Patsakis, Alexandros Asthenidis, and Abraham Chatzidimitriou carried out a case study with a malicious app on Facebook.<sup>6</sup> The app ostensibly was a slide show of dog pictures, but it also collected information about users' systems including IP address, browser version, operating system version, and open ports. Although the app only profiled users, it could have collected friend lists and sent messages to them, or executed arbitrary code.

Elias Athanasopoulos and colleagues examined ways to turn a social network into a botnet, demonstrating a proof-of-concept malicious Facebook app.<sup>7</sup> When a user activated the application, it displayed an image but also embedded hidden frames with inline images hosted at a designated target. Each time the user clicked within the app, it fetched the inline images without the target's awareness. The experiment suggests that an adversary taking full advantage of popular social utilities could generate a high volume of distributed denial-of-service traffic toward a target.

## References

1. L. Greiner, "Hacking Social Networks," *netWorker*, Mar. 2009, pp. 9-11.
2. W. Luo et al., "An Analysis of Security in Social Networks," *Proc. 8th IEEE Int'l Conf. Dependable, Autonomic, and Secure Computing (DASC 09)*, IEEE CS Press, 2009, pp. 648-651.
3. T.N. Jagatic et al., "Social Phishing," *Comm. ACM*, Oct. 2007, pp. 94-100.
4. G. Brown et al., "Social Networks and Context-Aware Spam," *Proc. 2008 ACM Conf. Computer Supported Cooperative Work (CSCW 08)*, ACM Press, 2008, pp. 403-412.
5. A. Besmer et al., "Social Applications: Exploring a More Secure Framework," *Proc. 5th Symp. Usable Privacy and Security (SOUPS 09)*, ACM Press, 2009, article no. 2.
6. C. Patsakis, A. Asthenidis, and A. Chatzidimitriou, "Social Networks as an Attack Platform: Facebook Case Study," *Proc. 2009 8th Int'l Conf. Networks (ICN 09)*, IEEE CS Press, 2009, pp. 245-247.
7. E. Athanasopoulos et al., "Antisocial Networks: Turning a Social Network into a Botnet," *Proc. 11th Int'l Conf. Information Security (ISC 08)*, LNCS 5222, Springer, 2008, pp. 146-160.

Inbox system, which functions as a closed e-mail service.

The unique features in Facebook include the Wall and News Feed. The Wall serves as a virtual bulletin board for people to post notes, comments, or other feedback about a person or group. Friends can write on one another's wall, and groups have walls for their members to communicate. News Feed aggregates and streams information about friends' activities.

The Facebook Platform was started in 2007 to let software developers create applications (in PHP or Java) that run in the Facebook environment. Facebook currently includes more than 550,000 active apps. These are actually installed on the developer's server, not on the Facebook server. Facebook calls an app when a user requests the application URL. The app communicates with Facebook using the Facebook API (application programming interface) or Facebook Query Language (FQL, similar to SQL). It returns content to Facebook formatted by the Facebook Markup Language (FBML, similar to HTML), which Facebook in turn presents to the user's Web browser.

Apps can interact and integrate with core Facebook services. For example, apps can access a user's friends list—say, to send invitations—or post to a user's news feed. One important built-in Facebook application is Links, which manages a user's link collection. Users can share links to interesting objects, and these links also appear on users' profile pages and in their news feeds. Recently, Facebook went further to offer Like buttons for any website. If a Facebook user clicks a Like button, the system adds a link to that website to the user's activity stream, which friends can see in their news feed.

## SECURITY THREATS

With their rising popularity, Facebook and other online social networks will become even more attractive targets than before.

Social engineering is an obvious attack vector because of the implicit trust most users have in the social network environment. A hacker can use a compromised account to send malware-infected messages to the account holder's friends, many of whom will accept the message at face value. Another social engineering attack lures users to a phishing site designed to look like Facebook and with a similar URL, and tries to trick them into submitting personal data.

Facebook users are also attractive targets for spam, including fake Facebook invitations, news stories, or Like messages. Fake e-mail becomes even more effective if an attacker can steal personal information from users' accounts.

Malware is another growing problem. A well-known example is the Koobface worm and its many variants, which have targeted Facebook as well as other online social networks such as Myspace, Twitter, and hi5 for more than two years. Koobface (an anagram of Facebook) spreads through messages to the friends of users who have an infected system. The message includes a video link that purportedly directs recipients to download an update to Flash Player but instead downloads the worm. Part of the worm payload is a Trojan horse that joins the computer to a peer-to-peer botnet.

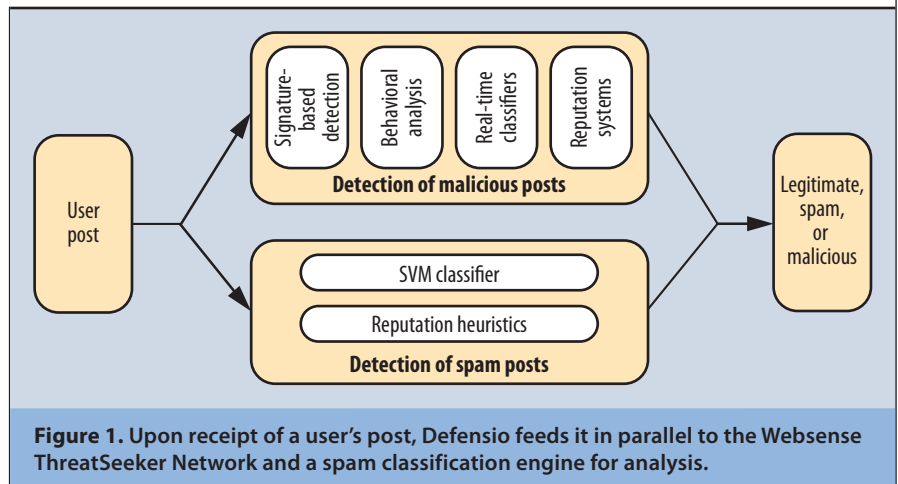
Malicious links are widely used for attacks, often taking users to a phishing site or drive-by download. Links can be shared in numerous ways in Facebook—for example, through messages, comments on a wall, shared news feed items, or the built-in Links application. A clickjacking worm has exploited the Like feature to spread such links: users receive messages with various subject lines that entice them to click a link; the link leads to a blank page with a hidden inline frame that publishes the initial message on their Facebook page, giving the appearance that they like the malicious link.

Facebook is quick to respond to suspicious or malicious links discovered by users or security companies and reportedly shares phishing and blacklist data with companies such as McAfee, MarkMonitor, and Microsoft. It also claims automated systems proactively detect and flag accounts with anomalous activity like sending many messages in a short time or messages with known bad links.

## DEFENSIO OVERVIEW

Defensio ([www.facebook.com/apps/application.php?id=177000755670](http://www.facebook.com/apps/application.php?id=177000755670)) is a Facebook application from Websense that monitors posts in a user's profile and determines whether they are legitimate, spam, or malicious (malware). In our study, we used the app to analyze only those posts that contained URLs. Although malicious links are clearly not the only threat to Facebook users, they are helpful in understanding the network's overall risk exposure.

To write a Facebook application, a developer registers with Facebook to access the Facebook API, which enables the app to read/write data from/to Facebook. In addition, Facebook provides an authentication mechanism that lets apps access the general information in users' profiles. It does not provide apps with access to users' private information; further, most apps require users' consent to access their data. When the user installs the app and allows it to



access his data, the app registers the user with Defensio and provides him with a Defensio key with the Defensio API.

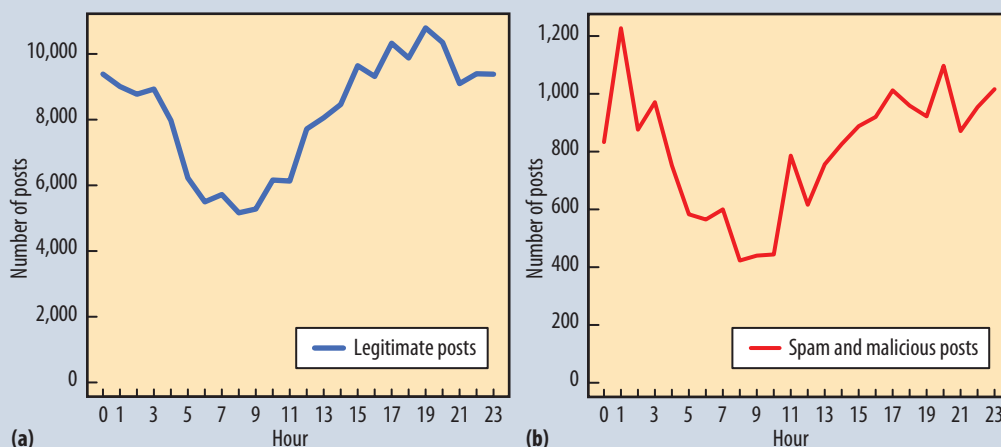
The app starts monitoring posts in the user's profile. It adds these to a *stream queue* and sends them in batches to Defensio for classification. The app associates each post with a Defensio user key to keep track of the recipient. After determining the status of the post, Defensio sends it to the *pending queue* to await user action. Users can request Defensio to immediately delete posts it classifies as spam or malicious, or they can request a notification e-mail and manually delete them.

Upon receipt of a user's post, Defensio feeds it in parallel to the ThreatSeeker Network ([www.websense.com/content/ThreatSeeker.aspx](http://www.websense.com/content/ThreatSeeker.aspx)), Websense's proprietary system for detecting malicious URLs, and a spam classification engine,<sup>3</sup> as Figure 1 shows. ThreatSeeker analyzes URLs in posts using a combination of signature-based detection, behavioral analysis of Web components, real-time content classifiers, and reputation systems, and accordingly flags those it determines to be malicious. The spam classification engine extracts the text from the post and runs it through a support vector machine (SVM) classifier, which assigns a score to the text. In parallel, the engine uses reputation heuristic rules to assign a score to the sender's identity. The engine then calculates a weighted average of the SVM and reputation scores and, based on this value, categorizes the post as either spam or ham (legitimate).

## RESULTS AND EVALUATION

Using Defensio data logs, we collected all Facebook posts containing a URL during a 21-day period, 22 June to 12 July 2010. These 502,624 posts were submitted by more than 25,000 users from 19 different countries. Each post had a timestamp indicating the date and time it was posted. In addition, Defensio had classified every post as legitimate, spam, or malicious. Our goal was not to evaluate the accuracy of Defensio's detection scheme but to





**Figure 2.** Facebook posts containing URLs per hour over 21 days: (a) legitimate; (b) spam and malicious.

survey the temporal and network-level properties of those posts containing URLs that Defensio had determined to be malicious or spam.

**Table 1. Network properties of Facebook posts.**

| Posts      | Unique hosts | IP addresses | IP blocks | ASNs  | Hosting countries |
|------------|--------------|--------------|-----------|-------|-------------------|
| All        | 11,352       | 6,552        | 2,931     | 1,588 | 78                |
| Legitimate | 10,393       | 6,256        | 2,828     | 1,541 | 74                |
| Spam       | 1,049        | 507          | 362       | 243   | 37                |
| Malicious  | 156          | 127          | 104       | 74    | 24                |

### Temporal properties

Approximately 215,999 of the Facebook posts contained URLs, averaging 10,286 each day. Thus, approximately two out of five posts contained a URL. The vast majority of these posts, 91 percent, were legitimate. A significant portion of posts, 8.7 percent, were spam: 18,693 total posts, averaging 890 per day. Only 0.3 percent of posts—644, an average of 31 per day—were malicious.

Figure 2 shows the number of Facebook posts at each hour totaled over all 21 days. The volume of legitimate posts rose steadily during the day to peak between 18:00 and 19:00 PST, perhaps because people socialize the most

**Table 2. Top 10 domains and their frequency.**

| All posts            |           | Legitimate posts     |           | Spam posts        |           | Malicious posts              |           |
|----------------------|-----------|----------------------|-----------|-------------------|-----------|------------------------------|-----------|
| Host                 | Frequency | Host                 | Frequency | Host              | Frequency | Host                         | Frequency |
| apps.facebook.com    | 115,560   | apps.facebook.com    | 102,464   | apps.facebook.com | 13,094    | nobrain.dk                   | 103       |
| facebook.com         | 42,010    | facebook.com         | 41,749    | facebook.com      | 223       | mcdonaldsexposed.info        | 63        |
| youtube.com          | 9,952     | youtube.com          | 9,781     | youtube.com       | 166       | facebook.com                 | 38        |
| foursquare.com       | 706       | foursquare.com       | 673       | myspace.com       | 116       | giveaway-madness.com         | 36        |
| reddit.com           | 477       | reddit.com           | 476       | open.spotify.com  | 69        | clicklikebro.info            | 21        |
| p.ly                 | 288       | p.ly                 | 288       | foursquare.com    | 33        | chkths.info                  | 20        |
| myspace.com          | 242       | feedproxy.google.com | 211       | runkeeper.com     | 28        | video.mcdonalds-revealed.com | 17        |
| hotmail.com          | 222       | causes.com           | 208       | ahmad.ly          | 20        | truth.mcdonalds-revealed.com | 14        |
| flickr.com           | 213       | hotmail.com          | 206       | flickr.com        | 19        | www.mjacksonisalive.com      | 12        |
| feedproxy.google.com | 211       | maximumpc.com        | 203       | hotmail.com       | 16        | thecoolapps.com              | 11        |

**Table 3. Top 10 ASNs and their frequencies.**

| All posts |           | Legitimate posts |           | Spam posts |           | Malicious posts |           |
|-----------|-----------|------------------|-----------|------------|-----------|-----------------|-----------|
| ASN       | Frequency | ASN              | Frequency | ASN        | Frequency | ASN             | Frequency |
| AS32934   | 148,036   | AS32934          | 135,731   | AS32934    | 12,272    | AS30736         | 104       |
| AS15169   | 10,575    | AS15169          | 10,383    | AS15169    | 181       | AS25653         | 63        |
| AS14618   | 1,683     | AS14618          | 1,633     | AS33739    | 101       | AS26347         | 53        |
| AS33070   | 927       | AS33070          | 870       | AS43650    | 74        | AS23522         | 36        |
| AS21844   | 924       | AS21844          | 844       | AS26496    | 52        | AS26496         | 36        |
| AS26496   | 726       | AS26496          | 638       | AS33070    | 52        | AS32934         | 33        |
| AS36351   | 638       | AS36351          | 623       | AS10913    | 50        | AS21844         | 32        |
| AS20940   | 398       | AS20940          | 398       | AS14618    | 49        | AS30058         | 20        |
| AS8075    | 395       | AS8075           | 371       | AS21844    | 48        | AS27458         | 15        |
| AS3561    | 349       | AS3561           | 340       | AS36024    | 26        | AS15169         | 11        |

after work. The pattern of spam and malicious posts looks roughly similar, rising during the day and early evening and then dipping in the early morning, but it also exhibits more irregularities. There was a sharp peak between 00:00 and 01:00 PST and a smaller peak between 19:00 and 20:00 PST. These irregularities might have occurred because spam and malicious posts are mostly planted by automated means, with the peaks representing bursts of activity by these programs.

### Network properties

For each URL extracted from the data, we resolved the IP address, autonomous system number (ASN), IP block, and country hosting the URL. To obtain the IP address-to-country mappings, we relied on MaxMind's geolocation database (<http://geolite.maxmind.com/download/geoip/database>), which uses regional Internet registries' whois information.

Table 1 summarizes the number of unique hosts, IP addresses, IP blocks, and ASNs, as well as the number of hosting countries. The values largely correspond to the proportionate volume of each category of URL. An unexpected revelation was that spam and malicious URLs were a small fraction of the total volume but were hosted in a disproportionately large number of countries, suggesting that malicious activities are geographically widespread.

Table 2 lists the top 10 domains and their frequencies (total number of appearances). Because the URLs were predominantly legitimate, the most frequently appearing domains in all posts were similar to those in legitimate posts. Most URLs in spam posts were hosted in the Facebook domain. In contrast, malicious links were hosted in various unusual domains.

Table 3 lists the top 10 ASNs and their frequencies. Because an ASN covers several IP addresses and IP blocks, we do not summarize the top IP addresses and IP blocks for

**Table 4. Top 10 ASNs not hosting any legitimate content.**

| ASN     | AS name                       | Country                |
|---------|-------------------------------|------------------------|
| AS6581  | BKCNET "SIA" IZZI             | Latvia                 |
| AS20597 | ELTEL-AS ELTEL.NET            | Russia                 |
| AS42560 | BA-GLOBALNET-AS               | Bosnia and Herzegovina |
| AS43134 | COMPLIFE-AS                   | Moldova                |
| AS19194 | JOVITA Sentris Network LLC    | US                     |
| AS34104 | GLOBAL-AS Iletisim Hizmetleri | Turkey                 |
| AS25751 | VCLK Valueclick Inc.          | US                     |
| AS29650 | HOSTING365-AS                 | Ireland                |
| AS30361 | SWIFTWILL2                    | US                     |
| AS50144 | LALIB-AS                      | Portugal               |

each URL category. Note that some ASNs hosted only spam or malicious content and no legitimate content. Table 4 lists the top 10 ASNs that only hosted malicious or spam content. AS6851, which tops the list, has been heavily linked with malicious activities, especially the Koobface worm.


Table 5 summarizes the top 10 hosting countries and their frequencies. Most of the URLs were hosted in the US, Denmark, Norway, the UK, and Canada. Checking these countries against the locations of the Defensio application users revealed that the majority of users were from the US, followed by Norway, Germany, the UK, and Canada, which explains why these countries top the list. All the URLs hosted in three countries—Latvia, Morocco, and Paraguay—appeared in malicious or spam posts.

**O**nline social networks are a convenient way to keep informed about activities, share messages and multimedia with family and friends, and meet new people with similar interests. At the same time, they expose users to numerous security threats.

Table 5. Top 10 hosting countries and their frequencies.

| All posts   |           | Legitimate posts |           | Spam posts  |           | Malicious posts |           |
|-------------|-----------|------------------|-----------|-------------|-----------|-----------------|-----------|
| Country     | Frequency | Country          | Frequency | Country     | Frequency | Country         | Frequency |
| US          | 187,340   | US               | 172,590   | US          | 14,344    | US              | 406       |
| Germany     | 1,215     | Germany          | 1,175     | Luxembourg  | 79        | Denmark         | 110       |
| Norway      | 989       | Norway           | 975       | Germany     | 33        | Malaysia        | 14        |
| UK          | 952       | UK               | 933       | Malaysia    | 28        | France          | 13        |
| Canada      | 395       | Canada           | 377       | Netherlands | 28        | Netherlands     | 10        |
| Netherlands | 320       | Namibia          | 282       | Canada      | 18        | Germany         | 7         |
| France      | 297       | France           | 266       | France      | 18        | UK              | 5         |
| Israel      | 239       | Israel           | 235       | UK          | 14        | Turkey          | 5         |
| Spain       | 221       | Spain            | 220       | Norway      | 14        | Latvia          | 4         |
| Denmark     | 208       | Italy            | 187       | Hong Kong   | 10        | Austria         | 3         |

Our study suggests that Facebook users can expect a significant chance (9 percent) of encountering spam posts and a much lower but not negligible chance (0.3 percent) of coming across malicious links. The study also found the domains in spam posts to be mostly commonplace while those in malicious links tend to be unusual. Not unexpectedly, links in spam and malicious posts appear to be primarily hosted in some of the most technologically advanced western countries, but a few smaller countries, such as Latvia, host a substantial proportion of malicious content and little or no legitimate content.

Malicious links are an important risk indicator but do not portray the entire threat landscape. Much more research is needed to gain a better grasp of the true extent and nature of security threats in online social networks, especially social engineering and malware. 

## Acknowledgment

The authors thank Websense Inc. for providing access to the Defensio data.

## References

1. N. Bilton, "Researcher Releases Facebook Profile Data," *The New York Times*, 28 July 2010; <http://bits.blogs.nytimes.com/2010/07/28/100-million-facebook-ids-compiled-online>.
2. S. Milgram, "The Small-World Problem," *Psychology Today*, May 1967, pp. 61-67.
3. S. Abu-Nimeh and T. Chen, "Proliferation and Detection of Blog Spam," *IEEE Security & Privacy*, Sept./Oct. 2010, pp. 42-47.

**Saeed Abu-Nimeh** is a senior researcher at Damballa Inc., a network security company based in Atlanta, Georgia. His research interests include Web security, spam and phishing detection, and machine learning. Abu-Nimeh received a PhD in computer science from Southern Methodist University. Contact him at [sabunimeh@damballa.com](mailto:sabunimeh@damballa.com).

**Thomas M. Chen** is a professor of networking in the School of Engineering at Swansea University, Wales. His research interests include Web filtering, Web classification, network traffic classification, smart grid security, privacy, cybercrime, and malware. Chen received a PhD in electrical engineering from the University of California, Berkeley. He is a senior member of IEEE. Contact him at [t.m.chen@swansea.ac.uk](mailto:t.m.chen@swansea.ac.uk).

**Omar Alzubi** is a PhD student in the School of Engineering at Swansea University. His research interests include machine learning for detecting Web threats, social network spam, and elliptic curve cryptography. Alzubi received an MS in computer and network security from the New York Institute of Technology. Contact him at [omar\\_alzubi@hotmail.com](mailto:omar_alzubi@hotmail.com).



Selected CS articles and columns are available for free at <http://ComputingNow.computer.org>.

Innovative Technology for Computer Professionals  
**Computer**

NEXT ISSUE  
**SOFTWARE  
ENGINEERING**