4th IEEE International Conference on Digital Ecosystems and Technologies (IEEE DEST 2010)


# Definition of Spam 2.0: New Spamming Boom

Pedram Hayati, Vidyasagar Potdar, Alex Talevski
Anti-Spam Research Lab
Digital Ecosystem and Business Intelligence Institute
Curtin University, Perth, Western Australia
{p.hayati,v.potdar, a.talevski}@curtin.edu.au

Nazanin Firoozeh, Saeed Sarenche, Elham A. Yeganeh
Institute for Advanced Studies in Basic Sciences
Zanjan, Iran
{n_firoozeh, sarenche, yeganeh}@iasbs.ac.ir

*Abstract*— **The most widely recognized form of spam is e-mail spam, however the term "spam" is used to describe similar abuses in other media and mediums. Spam 2.0 (or Web 2.0 Spam) is refereed to as spam content that is hosted on online Web 2.0 applications. In this paper: we provide a definition of Spam 2.0, identify and explain different entities within Spam 2.0, discuss new difficulties associated with Spam 2.0, outline its significance, and list possible countermeasure. The aim of this paper is to provide the reader with a complete understanding of this new form of spamming.**

***Keywords-spam, spam 2.0, web 2.0,***

## I. INTRODUCTION

Spamming – the act of spreading unsolicited and unrelated content – has been observed in several different domains such as email, instant messaging, web pages, Internet Telephony, etc [1-4].

Web 2.0 is commonly associated with web applications that facilitate interactive information sharing, interoperability, user-cantered design and collaboration on the World Wide Web [5]. The read/write Web 2.0 concepts are the backbone of the vast majority of web services that we use today. In contrast to non-interactive websites, Web 2.0 promotes an increasing emphasis on human collaboration through an architecture for participation that encourages users to add value to web applications as they use them. Today, Web 2.0 functions are commonly found in web-based communities, applications, social-networking sites, media-sharing sites, wikis, blogs, mashups, and folksonomies. They are widely provided by government, public, private and personal entities [5].

Spam is the abuse of electronic messaging systems to send unsolicited messages in bulk indiscriminately [6]. While the most widely recognized form of spam is e-mail spam, the term is applied to similar abuses in other media and mediums.

Spam 2.0 (or Web 2,0 Spam) is refereed to as spam content that is hosted on online Web 2.0 applications. Such spam differs from traditional spamming form in the following ways;

- it is targeted at Web 2.0 applications,
- current countermeasure are not suitable to detect and prevent spam 2.0 [7],
- spam 2.0 spreads through legitimate websites such as government, universities, personal homepages etc.
- it can be automatically distributed to as many Web 2.0 sites as possible through the use of automate agents [8]

The main problems with spam 2.0 are:

- get undeserved high ranking for spammer campaign in search engine results hence the low quality content get higher indexing position than good quality one,
- damage the reputation of legitimate website. A website that is infiltrated by spam 2.0 losses its attention of genuine users,
- waste valuable resources such as network bandwidth, memory space, etc,
- trick users, and damage popularity of systems.

This gives rise to significant socio-economic issues such as: 1) direct and indirect costs associated with the management of Spam 2.0, 2) reduction of Internet quality of service, 3) proliferation of Internet scams, viruses, trojans and malware.

There is no comprehensive statistic on amount of Spam 2.0 however, Live Spam Zeitgeist shows that amount of comment spam (comment spam is a part of spam 2.0) has doubled in 2009 as compared to 2008 [9].

In this paper: we provide a definition of Spam 2.0, identify and explain different entities within Spam 2.0, discuss new difficulties associated with Spam 2.0, outline its significance, and list possible countermeasure. The aim of this paper is to provide the reader with a complete understanding of Spam 2.0.

## II. DEFINITION OF SPAM 2.0

"Spam 2.0 is defined as propagation of unsolicited, anonymous, mass content to infiltrate legitimate web 2.0 applications"

The key element in Spam 2.0 definition is the distribution of spam content through legitimate websites. It differentiates this form of spamming from other types like email spam & web spam [1, 3]. Previously spammers use media such as email, instant messengers, Internet Telephony etc to spread spam hence such media serves as a communication medium between spammer and genuine users. However, spam 2.0 acts differently which uses legitimate web 2.0 applications to host spam. In spam 2.0, spammers no longer host their own email/web servers. Instead, they post spam on legitimate websites.

Legitimate websites here refer to those genuine website used by users such as governmental, universities, companies, homepages websites etc. Spam 2.0 infiltrates legitimate websites by posting/hosting on them. Examples of that include a promotional comment in blogs, a fake user profile, an unsolicited link in social bookmaking website etc.

## III. SPAM 2.0 SIGNIFICANCE

Spam 2.0 offers a far more attractive proposition for spammers as compared to traditional spam specifically email spam. Web 2.0 applications can be discovered through a simple search engine query that contains domain keywords and a web 2.0 application (e.g. "photography forum" for web 2.0 forum groups that are interested in photographs). Email addresses are procured in a similar fashion except email address details are commonly tightly controlled online and are far more difficult to source.

There are even efforts where false email addresses are hosted online to poison email spambot databases. In contrast, web 2.0 applications actively try to promote their existence so that web users are able to easily find them and contribute. Such applications rely upon relatively anonymous social network users to freely interact online.

Spammers can discover web 2.0 applications and use automated tools to distribute spam information that is targeted at a demographic of their choice with very little resistance. A single spam 2.0 attack may reach many targeted and domain specific users whereas a single email message would only potentially reach one random individual if the email address is real and it is not stopped by today's effective email spam filters.

Furthermore, once an individual discovers an email message that has bypassed their filters they are able to delete it. Messages online typically cannot be deleted by regular users and persist until an administrator deals with them often impacting many users in the meantime. Popular online discussion board rarely have more than one administrator for every one thousand users and a spam post may be overlooked and will persist online for extended periods.

Spam 2.0 posts also have a parasitic nature. They may exist on legitimate and often official websites. If such information persists, the trust in such pages is diminished, spam is effectively promoted by trusted sources, many users can be mislead or lead to scams and computer malware and such legitimate sites may be blacklisted which then deprives all others of legitimate content. As a result of the success and impact rates of spam 2.0, it is far more popular amongst spammers and has far greater negative socio-economic impact. Although other research has not confirmed the cost of web 2.0 spam, we believe it far exceeds email spam [10].

## IV. KEY ENTITIES IN SPAM 2.0

Figure 1 illustrates key entity involve in spam 2.0.

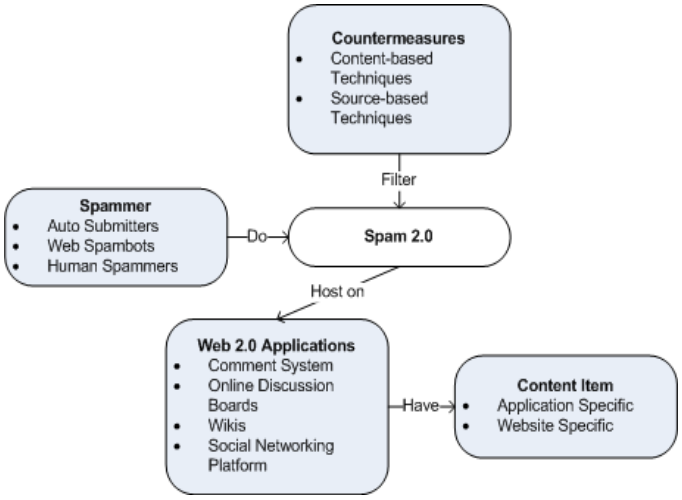


Figure 1. Spam 2.0 key entities

In following sections we discuss about each of above entities. We begin our discussion with content items inside each web 2.0 applications that targeted by spam 2.0.

Content item refers to discrete objects inside each web application that can be infiltrated by spammers. Content item can be *application specific* our *website specific*. Former, is objects that are exist in web applications such as comment systems, online discussion boards, wikis, etc (Table 1) while later are specific for website such as Amazon, Facebook, etc. Examples of content items include a post in an online discussion board, a product review in Amazon website, a profile in an online community website. We discuss about content item in each web 2.0 application that is compromised by spammer in following sections.

TABLE I. CONTENT ITEMS IN WEB 2.0 APPLICATIONS.

| | Comment | Forum | Wiki | Social networking |
|---|---|---|---|---|
| **Features** | comment content, trackback | post, poll, messaging, profile, attachment, signature | article content, tag, reference | profile, comment, tag, messaging, testimonial, attachment, |

### A. Comment systems

To date, comment systems are part of each and every web application in order to get users feedback on particular content. Examples of such systems include comments in blogs, product reviews, testimonials etc. Spammers use this opportunity to post promotional comments such as links to websites; product advertisement etc with an aim to drive traffic to certain sites or create a sale (Figure 2).


Figure 2. Example of spam 2.0 in comment system

## B. *Online discussion boards*

As shown in Table 1, Spammers can take advantages of forum content items such as post, poll, personal message, profile, attachment and signature for distributing spam content. A post is messages that created by registered users and contain text, image and date and time of sending message. A personal message is also a private message that users send to each other using forums. Spammers can register users in forums to spread spam content in different threads or send it as personal messages to other user. Spam content may contain links and unrelated keywords to spammers' campaigns. Spammers can leverage file-upload facility in online discussion board to attach unrelated images, documents, video, etc. files to their posts. Apart from that, spammers can modify their online discussion profile to publish spam content. Profiles may contain optional field for user interest, signature, homepage, etc. such field can be used by spammers to spread spam links or spam keywords. Figure 3 shows an example of spam 2.0 in an online discussion board.

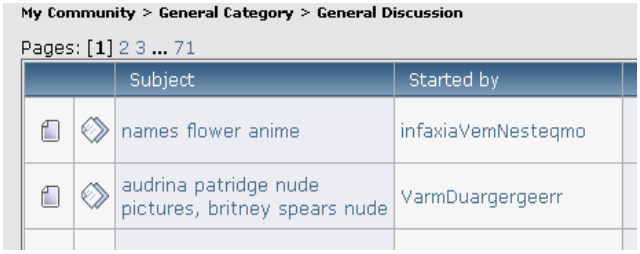


Figure 3. Example of spam 2.0 in an online discussion board

## C. *Wikis*

Wikis provide broader level of content contributions to web users. Inside wikis users can create/modify any number of webpages. However spammers are not omitted from such lists. They infiltrate wiki webpages with unrelated materials such as links or citation for content of the page. Particularly spammers try to spread out their link in wiki pages for search engine ranking rather than attracting users. Figure 4 shows examples of spam 2.0 in Wikipedia. The source of the citation link has been manipulated by spammers to the different source.


Figure 4. Examples of spam 2.0 in Wikipedia. The source of the citation link has been manipulated by spammers to the different source.

## D. *Social networking platforms*

Social bookmaking website where user can share videos and contribute to others submitted video, online communities where users can make a virtual identity and communicate with others, etc are few example of social networking platforms. Social networking platform provide different facilities for content sharing and communication Such as private and public messaging, comment system, interests, etc. However each of such facilities (Table 1) can be compromised by spammers to host spam like creating fake user profiles, sending private spam message, spam commenting on users profile, posting unsolicited videos, making promotional links etc. Hence such platform has more risk of being infiltrated by spam 2.0 and much harder to manage. Figure 5 shows an example of spam 2.0 in an online community. Spam 2.0 hosted on genuine user profile as a kind of comment.


Figure 5. Example of spam 2.0 in an online community

# V. HOW SPAM 2.0 WORKS

Currently spam 2.0 is propagated by techniques which include but may not be limited to *auto submitters, web spambots* or humans manual spamming.

Auto submitter is a kind of automated technique use by spammer to distribute spam to many website in fast and short amount of time. Such technique can be in form of tool that runs locally or it can be as service. They target online input forms such as contact forms, comment forms, etc for multiple consecutive submission.

Another automated but more sophisticated technique in spam 2.0 is web robots called Web spambots which is type of web robot [8]. Web robots, also called as Internet robots, are automated agents or scripts that perform specific tasks over the web in automatic fashion [11]. They execute structured repeating tasks faster than human can do. Since the web grows in size and value enormously, web robots play an essential role in the Internet community. Web robot is a double-edge sword: human assistance or a threat to web application [12]. Web spambot which is designed to facilitate sending spam is in threat edge. The early version of web spambot are email spambots that harvest email addresses from webpages, mailing lists, Directory databases, and chat rooms in order to send large number of unwanted emails [[13]. Email spambot can send many emails with efficiently and reliability. However, As Web 2.0 opens an avenue for interacting with people, spammers move toward misusing such capacities inside Web 2.0. Hence, new version of spambot called Web spambots has been developed to infiltrated web 2.0 applications. Web spambot are intended submitting spam content to web 2.0 application such as online discussion boards, wikis, commenting systems, and social networking platforms. Web spambots surf the Web to find web 2.0 applications and spread spam 2.0. They can target one specific web 2.0 application like Wikis as well as focus on one particular website such as MySpace [8].

Apart from web spambots, there maybe existence of real human spamming activities through hiring low-income labours [14]. This make spam detection task much more complicated than before since human's can intelligently bypass most of the anti-spam filters.

## VI. SPAM 2.0 COUNTERMEASURES

Although there has not been any comprehensive study aiming to combat spam 2.0 in general, literature have some attempts to prevent spam from web 2.0 applications. We group current countermeasures into two categories - Content-specific and Source-specific.

### *A. Content specific countermeasures*

Anti-spam solutions in this category attempt to discover patterns of spam 2.0 from spam content in web applications. Examples of such solutions are study by Zinman and Donath [15]. They develop a model classify spam profiles in Social Networking Services by extracting 40 features from profiles. Jindal and Liu [16] extract 36 features from reviews/opinion content inside review-gathering websites such as Amazon. Their approach train a classifier based those feature to classify opinion spam from legitimate opinion. A mechanism to identify video spammer in online social network such as Youtube has been proposed by Benevenuto et al. [17]. Their content-classifier use machine learning technique to classify video spammers. However, result of above studies as authors themselves mentioned are not satisfactory enough for spam classification.

### *B. Source specific countermeasures*

Solutions in this category target source of spam 2.0 problem i.e. identifying or preventing automated tools/web spambots. However majority of such solutions are design for targeting web robot detection and prevention in general rather than web spambots. Examples of source specific solutions are as follow. Tan et al. [13] propose a web robot session identification method based on their navigational patterns. The main assumption in their proposed system is that web robot navigational patterns such as session length and set of visited webpages (width and depth of visited webpages) are different from those of humans. The aim of their study is on unknown and camouflaged web robots and web crawlers. Park et al. [18] provide a method for malicious web robot detection based on types of requests for web objects (e.g. Cascading Style Sheet files, image files) and existence of mouse/keyboard activity. However, both above-mentioned studies did not focus on spambot detection in web 2.0 applications.

User-agent is a filed inside a HTTP Request Header send from web browser to the web server that identifies client application and it needs to be declared by web robot [11]. Hence by looking inside this field it is possible to block/allow specific web robot access to the website. However, spambots hide or fake their identity to other names [13]. So, it gets more complicate to detect spambot.

Some solutions try to slow down spambot activity but can not stop them. Hashcash is a technique use to slow down flow of automated-requests by increasing cost of submission in client side [19]. Sender of submission has to calculate a stamp which is difficult and time-consuming but comparatively cheap and fast for receiver to verify.

Completely Automated Public Turing test to tell Computers and Human Apart (CAPTCHA) is the most popular anti-robot technique adopted by many websites. It is a challenge response technique usually in format of a distorted image of letters and numbers [20]. Users are asked to infer a CAPTCHA image and type its letters in a form. On the other hand web-robots cannot infer CAPTCHA hence it prevents them from entering to the website.

Our works on web spambot detection [7, 21] proposes anti-spambot technique based web spambot web usage behaviour. The main assumption of our proposed method is that web spambots' web usage behaviour is intrinsically different from human ones. By intrinsically aggregating web usage data we formulate web usage data into 3 features sets to extract web spambot behaviour. Our promising result shows that our system is capable of classifying web spambot from humans in web 2.0 applications.

## VII. CONCLUSION

In this paper we propose a definition for new type of spamming boom called spam 2.0. Spam 2.0 is defined as propagation of unsolicited, anonymous, mass content to infiltrate legitimate web 2.0 applications. Spam 2.0 is different from traditional spamming technique since it has parasitic nature and it hosted on legitimate web applications. One of the most popular tools used by spammers is web spambots. web spambots can crawl the web, find web applications and spread spam 2.0. Current spam 2.0 countermeasures are whether look for spam pattern inside content or identify and prevent web spambots from entering to the website. The area of research in

this filed is quite young and current solutions are not effective enough.

## REFERENCES


[1] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, "A Bayesian Approach to Filtering Junk E-Mail," in *Learning for Text Categorization: Papers from the 1998 Workshop*, 1998.

[2] A. Cournane and R. Hunt, "An analysis of the tools used for the generation and prevention of spam," *Computers & Security,* vol. 23, pp. 154-166, 2004.

[3] Z. Gyongyi and H. Garcia-Molina, "Web spam taxonomy," in *Proceedings of the 1st International Workshop on Adversarial Information Retrieval on the Web*, Chiba, Japan, 2005.

[4] P. So Young, K. Jeong Tae, and K. Shin Gak, "Analysis of applicability of traditional spam regulations to VoIP spam," in *Advanced Communication Technology, 2006. ICACT 2006. The 8th International Conference*, 2006, pp. 3 pp.-1217.

[5] T. O'Reilly, "What Is Web 2.0," in *http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html:* O'Reilly Network, 2005.

[6] B. Whitworth and E. Whitworth, "Spam and the social-technical gap," *Computer,* vol. 37, pp. 38-45, 2004.

[7] P. Hayati, K. Chai, A. Talevski, and V. Potdar, "Behaviour-Based Web Spambot Detection by Utilising Action Time and Action Frequency," in *The 2010 International Conference on Computational Science and Applications (ICCSA 2010)*, Fukuoka, Japan, 2010.

[8] P. Hayati, K. Chai, V. Potdar, and A. Talevski, "HoneySpam 2.0: Profiling Web Spambot Behaviour," in *12th International Conference on Principles of Practise in Multi-Agent Systems*, Nagoya, Japan, 2009, pp. 335-344.

[9] Live-Spam-Zeitgeist, "Some Stats, Akismet," [Accessed online by May 2009] http://akismet.com/stats/, 2009.

[10] F. Ridzuan, V. Potdar, and A. Talevski, "Key Parameters in Identifying Cost of Spam 2.0," in *24th IEEE International Conference on Advanced Information Networking and Applications (AINA 2010)*, Perth, Western Australia, 2010.

[11] M. Koster, "Guidelines for Robot Writers [Accessed Online Dec 2009]," in *http://www.robotstxt.org/guidelines.html*, 1993.

[12] M. Koster, "Robots in the Web: threat or treat? [Accessed Online Dec 2009]," in *http://www.robotstxt.org/threat-or-treat.html*, 1995.

[13] P.-N. Tan and V. Kumar, "Discovery of Web Robot Sessions Based on their Navigational Patterns," *Data Mining and Knowledge Discovery,* vol. 6, pp. 9-35, 2002.

[14] Workathome, "Work from home online ad placing work pay per posting," [Accessed: 3 Aug 09] http://www.workathomeforum.in/online-adplacing-homejob.htm [Archived: http://debii.curtin.edu.au/~pedram/archive/online-adplacing-homejob.htm], 2009.

[15] A. Zinman and J. Donath, "Is Britney Spears spam," in *Fourth Conference on Email and Anti-Spam* Mountain View, California, 2007.

[16] J. Nitin and L. Bing, "Opinion spam and analysis," in *Proceedings of the international conference on Web search and web data mining* Palo Alto, California, USA: ACM, 2008.

[17] F. Benevenuto, T. Rodrigues, V. Almeida, J. Almeida, C. Zhang, and K. Ross, "Identifying Video Spammers in Online Social Networks," in *AIRWeb '08* Beijing, China, 2008.

[18] K. Park, V. S. Pai, K.-W. Lee, and S. Calo, "Securing Web Service by Automatic Robot Detection," *USENIX 2006 Annual Technical Conference Refereed Paper,* 2006.

[19] D. Mertz, "Charming Python: Beat spam using hashcash," [Accessed: 3 Aug 09] http://www.ibm.com/developerworks/linux/library/l-hashcash.html, 2004.

[20] L. von Ahn, M. Blum, N. Hopper, and J. Langford, "CAPTCHA: Using Hard AI Problems for Security," in *Advances in Cryptology — EUROCRYPT 2003*, 2003, pp. 646-646.

[21] P. Hayati, V. Potdar, K. Chai, and A. Talevski, "Web Spambot Detection Based on Web Navigation Behaviour," in *24th IEEE International Conference on Advanced Information Networking and Applications (AINA 2010)*, Perth, Western Australia, 2010.