Chapter 8
Spam 2.0 State of the Art

Pedram Hayati
Curtin University, Australia

Vidyasagar Potdar
Curtin University, Australia

ABSTRACT

Spam 2.0 is defined as the propagation of unsolicited, anonymous, mass content to infiltrate legitimate Web 2.0 applications. A fake eye-catching profile in social networking websites, a promotional review, a response to a thread in online forums with unsolicited content, or a manipulated Wiki page are examples of Spam 2.0. In this paper, the authors provide a comprehensive survey of the state-of-the-art, detection-based, prevention-based and early-detection-based Spam 2.0 filtering methods.

INTRODUCTION

Web 2.0 is commonly associated with web applications that facilitate interactive information sharing and collaboration on the Internet. Web 2.0 promotes an increasing emphasis on human collaboration that encourages users to add value to web applications as they use them, such as social-networking sites, media-sharing sites, wikis, blogs, etc. Spam abuses such systems by sending unsolicited messages in bulk. The intentions of such spam are to misinform users (scams), generate traffic, generate sales, and spread spyware or malware. While the most widely recognized form of spam is e-mail spam, the term is applied to similar abuses in other media. Figure 1 illustrates different spam types for different platforms.

"Spam 2.0 is defined as the propagation of unsolicited, anonymous, mass content to infiltrate legitimate Web 2.0 applications" (Hayati et al., 2010a). A fake eye-catching profile in social networking websites, a promotional review, a response to a thread in online forums with unsolicited content and a manipulated Wiki page, etc., are some examples of a new form of spam on the web that is referred to as Spam 2.0 (Hayati et al., 2010a).

In the context of Spam 2.0, web spambots (or *spambots*) are used to automatically and repetitively crawl the web, find vulnerable Web 2.0 applications and send bulk messages indiscriminately. Typically, spammers, scammers and hackers collaborate to attack networks, destroy cyber infrastructure, hijack computers, spy on private/

DOI: 10.4018/978-1-4666-4006-1.ch008

Figure 1. Evolution and adaptability of spam once new media emerge

confidential data, obtain privileged information and spread spam.

During the past decade, the Internet has accumulated a significant amount of Spam 2.0 that is continually growing. Recent statistics show that up to 87.6 million Web 2.0 applications are infected with Spam 2.0 each year (Sophos, 2009). Furthermore, current research indicates that 75%+ of pings coming from blogs to search engines for the update of information are spam (Kolari, Java, & Finin, 2006) and the amount of comment spam in 2009 doubled that of 2008 (Akismet, 2011). Such an overwhelming amount of spam is seriously degrading the quality of information on the Internet (Chai, Potdar, &Dillon, 2009).

Furthermore, Spam 2.0 offers a far more attractive proposition for spammers than does email spam. Web 2.0 applications can be easily discovered through a simple search engine query that contains domain keywords and a Web 2.0 application name. Spammers can discover Web 2.0 applications and use automated spambots 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 if it is not blocked by today's effective email spam filters.

In addition, once an individual discovers an email message that has bypassed their filters, they are able to delete it. However, online messages typically cannot be deleted by regular users and persist until an administrator deals with them, thus often impacting on many users in the meantime. Popular forums rarely have more than one administrator for every one thousand users and a spam post may be overlooked and persist online for extended periods. Spam 2.0 posts also have a parasitic nature. They often exist on "legitimate" and often official websites.

It is very important to understand the term "legitimate", in the context of Spam 2.0. It refers to Web 2.0 applications or websites that are designed with a genuine purpose, i.e., to serve web users (e.g., YouTube™ for video sharing, Facebook™ for social networking, etc.), which is then exploited by spammers for spamming, by relying on interactive content generation facility offered by the Web 2.0 platforms.

Spam 2.0 infiltrates through such "legitimate" websites by posting spam content on them. Figure 2, Figure 3, and Figure 4 illustrates few examples of Spam 2.0 in legitimate websites: blog comments, online communities and forum posts. This is the key differentiator between Spam 2.0 and other spam types (Hayati et al., 2010a). Spammers no

Figure 2. A comment spam