Data Mining for Obtaining Secure E-Mail Communications

Mª Dolores del Castillo
Instituto de Automática Industrial (CSIC), Spain

Ángel Iglesias
Instituto de Automática Industrial (CSIC), Spain

José Ignacio Serrano
Instituto de Automática Industrial (CSIC), Spain

INTRODUCTION

Email is now an indispensable communication tool and its use is continually growing. This growth brings with it an increase in the number of electronic threats that can be classified into five categories according to their inner behavior: virus, trojans, pharming, spam, and phishing. Viruses, trojans and pharming threats represent an attack to the user's computer while the focus of attack of spam and phishing threats is mainly the user, that is, these last two threats involve a kind of intellectual attack.

A virus is a small program that replicates itself and inserts copies into other executable code or documents using e-mails as a means of transport. Trojans can not replicate themselves and they are used to open a network port giving other users a means of controlling the infected computer. Other more dangerous trojans are called spy programs (spyware) which wait until users visit some websites and then capture all the keys typed and mouse movements and make screenshots to obtain information. Pharming is a technique used to redirect users to illegitimate websites. These three threats, in spite of being present in e-mails, can be solved by an anti virus program.

The next two threats need e-mail filters to be solved and this chapter focuses on them: spam and phishing. Spam consists on the massive sending of unsolicited commercial e-mail to a large number of recipients. Unlike legitimate commercial e-mail, spam is sent without the explicit permission of the recipients. Spammers obtain e-mail addresses by different ways such as guessing common names at known domains or searching addresses in web pages. A report from the Commission of European Communities ("Communication from", 2004) shows that more than 25 percent of all e-mail currently received is spam. More recent reliable data shows that spam represents 60-80 percent of e-mail volume. Spam is widely recognized as one of the most significant problems facing the Internet today.

Spam has evolved to a new and dangerous form known as 'phishing'. Phishing differs from spam in that it is generated by a criminal intent on stealing personal data for financial gain ("Spyware", 2007). Phishing is the term used to describe emails which trick recipients into revealing their personal or their company's confidential information such as social security and financial account numbers, account passwords and other identity or security information.

According to Anti-Phishing Working Group ("June Phishing", 2006) the number of phishing reports has increased from 20,109 in May 2006 to 28,571 in June 2006 and it is the most ever recorded. Phishing attacks increase despite of the efforts of e-mail filters. Although only 0.001 percent of e-mail sent is responded to, this percentage is enough to return on the investment and keep the phishing industry alive. Further research has estimated that the costs of these phishing attacks on consumers in 2003 ranged from $500 million to an amazing $2.4 billion.

The early phishing attempts consisted on a link to a website which looked like a legitimate website, but in fact was an illegitimate website. The website address usually was not a domain, but simply an IP address, and the e-mails were often very poorly written, with bad grammar and spelling, and little attention to detail. Needless to say that phishing attacks have evolved with more convincing content and became harder to recognize. While a non-professional appearance such as a spelling error, a dubious URL, or a non-secure website are sure signs of a fraudulent phishing website, the lack of these features can no longer be used as a sure sign of a legitimate site (Green, 2005).

It is hard to successfully obtain bibliographical information in the scientific and marketing literature about techniques that aim to avoid spam and electronic fraud. This could be due to the features of these security systems, which should not be revealed in public documents for security reasons. This lack of information prevents improvements of criminals' attacks because spammers/phishers just do not know the methods used to detect and eliminate their attacks. It is also necessary to emphasize that there is little available commercial technology that shows an actual and effective solution for users and businesses.

Spam and phishing filters process e-mail messages and then choose where these messages have to be delivered. These filters can deliver spurious messages to a defined folder in the user's mailbox or throw messages away.

BACKGROUND

Filtering can be classified into two categories, origin-based filtering or content-based filtering, depending on the part of the message chosen by the filter as the focus for deciding whether e-mail messages are valid or illegitimate (Cunningham, 2003). Origin-based filters analyse the source of the e-mail, i.e. the domain name and the address of the sender (Mertz, 2002) and check whether it belongs to white (Randazzese, 2004) or black (Pyzor, 2002) verification lists.

Content-based filters aim to classify the content of e-mails: text and links. Text classifiers automatically assign document to a set of predefined categories (legitimate and illegitimate). They are built by a general inductive process that learns, from a set of preclassified messages, the model of the categories of interest (Sebastiani, 2002). Textual content filters may be differentiated depending on the inductive method used for constructing the classifier in a variety of ways:

Rule-based filters. Rules obtained by an inductive rule learning method consist of a premise denoting the absence or presence of keywords in textual messages and a consequent that denotes the decision to classify the message under a category. Filters that use this kind of filtering assign scores to messages based on fulfilled rules. When a message's score reaches a defined threshold, it is flagged as illegitimate. There are several filters using rules (Sergeant, 2003).

Bayesian filters. First, the probabilities for each word conditioned to each category (legitimate and illegitimate) are computed by applying the Bayes theorem, and a vocabulary of words with their associated probabilities is created. The filter classifies a new text into a category by estimating the probability of the text for each possible category C_j, defined as $P(C_j \mid text) = P(C_j) \cdot \Pi_i P(w_i \mid C_j)$, where w_i represents each word contained in the text to be classified. Once these computations have been carried out, the Bayesian classifier assigns the text to the category that has the highest probability value. Filter shown in ("Understanding Phishing", 2006) uses this technique.

Memory-based filters. These classifiers do no build a declarative representation of the categories. E-mail messages are represented as feature or word vectors that are stored and matched with every new incoming message. These filters use e-mail comparison as their basis for analysis. Some examples of this kind of filter are included in (Daelemans, 2001). In (Cunningham, 2003), case-based reasoning techniques are used.

Other textual content filters. Some of the content-based approaches adopted do not fall under the previous categories. Of these, the most noteworthy are the ones based on support vector machines, neural networks, or genetic algorithms. SVMs are supervised learning methods that look, among all the n-dimensional hyperplanes that separate the positive from the negative messages training, the hyperplane that separates the positive from the negative by the widest margin (Sebastiani, 2002). Neural network classifiers are nets of input units representing the words of the message to be classified, output units representing categories, and weighted edges connecting units represent dependence relations that are learned by backpropagation (Vinther, 2002). Genetic algorithms represent textual messages as chromosomes of a population that evolves by copy, crossover and mutation operators to obtain the textual centroid or prototypical message of each category (Serrano, 2007).

Non textual content-based filters or link-based filters detect illegitimate schemes by examining the links embedded in e-mails. This filtering can be done by using white-link and black-link verification lists (GoDaddy, 2006) or by appearance analysis (NetCraft, 2004) looking for well known illegitimate features contained in the name of the links.