ZERO-HOUR, REAL-TIME COMPUTER VIRUS DEFENSE THROUGH COLLABORATIVE FILTERING

Vipul Ved Prakash, Founder and Chief Scientist, Cloudmark, Inc.
Adam J. O'Donnell, Senior Research Scientist, Cloudmark, Inc.

January 2006

ABSTRACT

Conventional anti-virus software relies on a staff of researchers to isolate and analyze viruses, identify them with a fingerprint, and then write and test code and rules to block them. This process takes up to 24 hours and often blindly blocks many legitimate messages with attached executable code. In contrast, Cloudmark's Global Threat Network™ uses a fingerprinting algorithm to identify each incoming message, combined with a reputation-based, trusted community of real-time users to identify malicious viruses. Using the Cloudmark Trust Evaluation System™ (TES), Cloudmark is able to let trustworthy, credible users identify viruses. Cloudmark's virus fingerprinting algorithm automates the time-intensive "reverse engineering" analysis of conventional technologies allowing its system to identify and squelch new worms and virus strains in real time. The Cloudmark technology is language-agnostic, format-agnostic, representation-agnostic, and protocol-agnostic — making it particularly suited to combat all forms of malicious content.

Individuals who use messaging services, such as e-mail, SMS, and MMS, are really good at discerning the difference between content that is spam and content that is legitimate. When users are pooled together and allowed to "vote" on which content is and is not spam, the quality of their opinions is astonishingly high. The concept of pooling individual opinions regarding a piece of data is known as collaborative filtering.

The actual process of collaborative filtering is complex, but intuitive, and can be broken down into several components: fingerprint generation, weighted voting, fingerprint acceptance, and filtration. The fingerprint-generation phase begins with the computation of a mathematical function on each e-mail message that is received by the client. The output of the function, known as the fingerprint, is a numerical value that can be created only from that exact e-mail and specifically-associated variations of that e-mail. By generating fingerprints that are flexible to small variations, Cloudmark makes it extremely difficult for spammers to change a slight amount of text in order to evade previously-generated fingerprints.

The fingerprints are submitted to a central database during the weighted voting period. A user votes on whether a message is spam by submitting the fingerprint with a flag indicating that they believe the content to be a piece of spam. The weight assigned to their vote is based upon the user's historical trustworthiness in correctly reporting spam. Therefore, if a user has a hard time determining if a piece of content is spam or not spam, their vote will ultimately count less. Once the number of weighted votes on a specific fingerprint reaches a predetermined threshold, the fingerprint is accepted as a verified indicator of spam. From that point forward, any user who submits that fingerprint will be informed that the community has determined the content to be spam, and the system will filter out the message.

The algorithm for weighting user votes, which determines the correct point at which to accept a fingerprint, and which handles disagreement inside the community regarding the "spamminess" of a message, is encoded in the Cloudmark Trust Evaluation System, or TES. Just as in a real-world human community, individuals who behave well by quickly and correctly identifying spam become trusted, and are rewarded by having their opinions weighted more heavily in the future.

The principle of community-based collaborative filtering was critical in the design of our anti-

spam system, known as the Cloudmark Global Threat Network[1].
The Cloudmark system is not limited to the collaborative filtering of e-mail. From the outset, the Cloudmark system has been designed to implement a general framework for filtration, based on collaborative opinion. The core of the classification system is language-agnostic, format-agnostic, representation-agnostic, and protocol-agnostic. The Cloudmark approach can be rather easily adapted to almost any messaging medium by enabling feedback and filter hooks into the medium.

ANTI-VIRUS

Conventional anti-virus software works on a basic principle. The software examines every file that comes into the machine and generates a unique fingerprint for each file. This fingerprint is then checked against a fingerprint database of known viruses. The generation of the fingerprints traditionally requires a human being in the loop. The staffs at anti-virus research companies first have to isolate individual samples of computer viruses while they are propagating in the wild, disassemble them to study their code, generate a fingerprint for the virus, pass the fingerprint through a QA process, and finally, distribute the fingerprint to the clients.

The problem inherent in this model is the latency introduced by, both, the extraction of a virus sample, and the human examination of virus code. Viruses are collected either by manually gathering samples from the wild, or by manually examining honeypots, or by machines solely dedicated to accepting and recording attacks for later analysis. The latency associated with the collection and analysis phase can be several hours and, at times, the examination effort alone can exceed 24 hours. Anti-virus security infrastructures were developed for very specific environmental conditions, and they worked extremely well for their time. However, the environment has changed considerably in the last two years. These days, all major security vulnerabilities in operating systems and popular software are rapidly turned into worms. Worms automatically exploit these vulnerabilities to spread themselves across the Internet. And they do this very, very rapidly.

In response, new technology that claims to be “zero-time” has appeared in security products. Most of this technology is heavy-handed, in that it stops viruses by blindly blocking anything that resembles executable code. While such technology is reasonably effective in deterring virus and worm epidemics, it comes at a cost. The collateral damage from blindly blocking certain content types leads to blockage of legitimate traffic. This method can actually exacerbate lost productivity since it requires immediate intervention at the client level to retrieve legitimate, and often, business-critical messages—not to mention the additional administrative burden this approach places on strained IT departments.

Cloudmark addresses the problem of rapid response and precise identification of viruses by leveraging its massive installed base of human malware reporters and by introducing a special fingerprinting algorithm for capturing executable code. Since any one of Cloudmark's many users may potentially be the first person to see an emergent e-mail worm, our installed base effectively becomes the world's largest network of honeypots for malicious e-mail content, dramatically reducing our virus sample collection time. Additionally, Cloudmark's executable fingerprinting algorithm automates the process of human examination of viruses by disassembling executable code in real time and generating fingerprints that uniquely and robustly identify worms, viruses, and strains thereof. This fingerprinting algorithm automates the "reverse engineering" process used by conventional anti-virus vendors, and successfully removes many of the obfuscations hackers place on binaries to evade detection. By automating this process, Cloudmark's technology essentially removes the arduous, human-driven task of malware analysis. In practice, the Cloudmark anti-virus engine is able to detect and filter viruses as soon as users click "block," which is well before they are even named by the security community. It is for this reason that in many trials, Cloudmark's anti-virus engine has proved superior to conventional anti-virus software, often stopping 10-20 times as many viruses as the leading vendor.

As stated earlier, the Cloudmark system uses content and bearer protocol-agnostic algorithms for the generation and validation of fingerprints for spam and malicious content. As long as Cloudmark can provide a means for users or organizations to generate feedback regarding what messages are and are not spam, it is possible to filter spam and viruses out of any messaging scheme.

THREATS OF THE FUTURE

Advanced societies have evolved a standard methodology for dealing with new threats. They come together and determine, through consensus, the nature of, and appropriate response to, the threat. The Cloudmark system is a generic codification of this process, where users can submit fingerprints for any form of content, including viruses and assorted forms of malware, then vote upon the disposition of the content. Due to its content-agnostic engineering, our technology can be extended to combat all forms of malicious content, ranging from today's virus threats to those which have not yet been designed.

REFERENCES

1. V. V. Prakash and A. O'Donnell. *Fighting spam with reputation systems*. Queue, 3(9):36–41, 2005.

For more information visit us at
www.cloudmark.com
128 King Street, 2nd Floor
San Francisco, CA 94107 USA
Ph: +1.415.543.1220
Fax: +1.415.543.1233