

COMBATING SPAM

Best Practices

March 2007

OVERVIEW

Spam, Spam, More Spam and Now Spyware, Fraud and Forgery

Spam used to be just annoying, but today its impact on an organization can be costly in many different ways. The rise in email fraud and forgery, a tactic often used in phishing attacks, can have serious consequences to a business and its customers. The most common example of this kind of email forgery is designed to steal private customer information, such as credit card information and other personal data. The cost to a business can be devastating.

It is estimated that spam has increased by more than 200% since 2005 and the sophistication of spammers continues to advance at an alarming pace. Bot-nets, networks of computers infected with malware used for spam distribution, are the preferred method for distributing spam. Spam content delivered in the form of images rather than text has increased to 40% of the spam while the average size of a spam message has increased from 1KB to 40KB. Additionally, there is a substantial increase in false negatives (spam email not being detected as spam), while spam detection rates remain unchanged.

“The number of phishing spoof sites reached an all time high of nearly 30,000 unique phishing URLs...and in the month of January 2007, the total number of brands hijacked increased to 135 with brokerages and international banks’ among the top targets.”

— Anti-phishing Working Group
January 2007 Report

The volume, complexity, and speed of attacks continue to increase even as organizations pour additional resources into fighting the problem. Increasing volumes of spam not only impact users and customers; it also puts tremendous stress on the messaging infrastructure. The cost for businesses to battle this onslaught of spam, which now constitutes 93% of Internet traffic, continues to rise.

The implications for these trends to businesses are:

- Increase in liability from fraud and forgery
- End-user perceptions that anti-spam scanning technology is getting less effective
- Infrastructure costs (throughput, storage, etc) are drastically increasing
- Continual investment in anti-spam scanning infrastructure is required
- A paradigm shift in boundary security is required

To effectively battle this problem, reduce liability risks and lower costs, organizations should evaluate their spam fighting best practices on an on-going basis. The spam problem must be addressed with a combination of an architectural approach and advanced technological solutions. Buying another spam scanner is not enough. A holistic approach to winning the spam battle is required.

“An average of 30 percent of an email servers capacity is taken by spammers trying to steal email addresses and other information stored in corporate directories.”

ARCHITECTURAL BEST PRACTICES

Gateway Bounce Management

Standard Internet security practice is to use gateway hosts at the edge of the network to be the point of contact between the public and the private network for network-based services, including email. This email gateway should perform as much of the spam filtering as possible to put the burden of handling bounces on the sender. If recipient addresses are validated at the edge of the network, bounces are handled by the sending (spammer) not the receiving host. This can substantially improve

overall server utilization, improve total message throughput, and lower infrastructure costs.

When a message is bounced, a delivery status notification (DSN) is generated in order to return the message as undeliverable. The machine receiving the delivery error (the machine that last had possession of the message, which is the sending host) is the host responsible for generating and delivering the DSN. The most common reason for bouncing a message is that the recipient address does not exist. If the receiving host were to accept the message without validating the recipient address, then the burden of handling the DSN is now on your organization. Because the sender's address on most spam is forged, the DSN is typically undeliverable. The DSN will queue for future delivery attempts if the host to which it needs to be sent is unreachable. When the DSN is not deliverable, it sits in the mail queue until it expires and is bounced to the local mail administrator account as undeliverable. As that message sits in a delivery queue waiting to be retried, it takes up valuable system resources. Also, if the message is delivered to a forged address, your organization is now responsible for mailing to an innocent victim which can affect your reputation. Therefore, handling DSN's is an expensive proposition and the burden should be put back on the sending (spammer) host.

Gateway Content Scanning

Since messages are first received from sending hosts at the network edge, this makes connecting IP address information available to spam filters. This is a critical point in the message flow to perform content scanning. IP reputation, RBLs (realtime blackhole lists), some anti-spam fingerprinting technologies, blacklists and whitelists depend on connecting IP information. This IP address information is only available at connection time, unless it is parsed from Received: headers in the message, which is much more computationally expensive. Therefore, it is important to take advantage of this IP information and perform content scanning at this critical point in the message flow.

Content Scanning Approaches

There are four basic approaches to content scanning: heuristic, probabilistic, community based fingerprinting, and network based fingerprinting.

Heuristic methods generally look at content, headers, and other markers defined in a message, apply a score to them and when the score reaches a certain level, the content is called spam. These were some of the first methods available to fight spam, but it is a continual arms race to stay ahead of the spammers.

Probabilistic methods such as Bayesian classification calculate statistical probabilities that a message is spam, based on a set of message features. They tend to be highly accurate when well trained and applied to an individual recipient's mail stream. When applied to an organization's mail stream the classifiers may be over trained and suffer from either too many false positives or a lack of classification accuracy or both.

Community-based fingerprinting is a method which uses spam samples from a large cohort of end users who report spam. A set of fingerprints that define a given spam campaign are generated and it is distributed in the form of a database for use in filters. Cloudmark is the maintainer of the Razor database for reporting spam samples and offers an anti-spam filter based on that database and other sources obtained commercially. The method is highly accurate with low false positives. The challenge is to keep the fingerprinting technologies up to date as spammers continually attempt to defeat them.

Network based fingerprinting is a method which collects samples of messages at major peering points on the Internet and applies statistical algorithms to identify spam campaigns and build signature databases of those messages. It has the benefit of also being very effective at identifying virus outbreaks based on the distribution patterns of the messages around the Internet. Commtouch's Recurrent Pattern Detection (RPD™) is one such system for catching spam.

Honey pots are a way to capture spam samples. Accounts that are never used but that receive spam are by definition catching only spam and make an excellent source for generating

fingerprints and training classifiers and most anti-spam technology providers use honey pots for their system.

All strategies are an arms race with the spammers who morph their content or sending behavior to avoid classification as spam. Fingerprinting technologies tend to be the most rapidly adaptable to new spamming methods and heuristic methods are the most exploited by spammers with their morphological changes.

End User Quarantine

In addition to performing as much of the scanning as possible at the email gateway, providing an end-user accessible quarantine reduces the administrative overhead of the anti-spam solution. The quantity of spam is such that a single administrative quarantine for all spam is impractical for actually finding a message that was misclassified. A single quarantine also generates more support calls, because the user, who suspects a legitimate message was caught, cannot check their own quarantine.

TECHNOLOGICAL BEST PRACTICES

A well thought out defense-in-depth approach employing a number of technologies is required to effectively fight today's sophisticated spam. These technologies should be deployed in the order below, which will reduce the computational load as much as possible, thus increasing message throughput and reducing infrastructure costs.

The following methods and technologies are recommended to more effectively fight spam:

- Set the Greet Pause feature in the Sendmail MTA
- Implement Connection Controls
- Use IP Reputation Services
- Put Sender Authentication into practice
- Perform aggressive Spam Scanning

All of these methods, which are part of Sendmail's Sentrion™ email gateway appliance and software solutions, can significantly reduce the spam that reaches the end-users' inboxes, while minimizing misclassified good mail, and lowering deployment and infrastructure maintenance costs.

Greet Pause

The greet pause feature of the Sendmail MTA fights hit and run spammers. Spammers often attempt to get as much data to the client as soon as possible by sending their SMTP commands without waiting for the receiving MTA to acknowledge the commands they sent. By enforcing even a 200-millisecond delay before the SMTP greeting, a substantial fraction of mail from nefarious senders is eliminated.

Connection Controls

Rate limiting may reduce the amount of spam received from bulk senders. Limits on the total number of concurrent connections a sender may open, the rate of envelopes (separate messages) a sender may send, and the number of recipients allowed on a single message, when combined removes a significant fraction of email from nefarious senders. This is accomplished merely by issuing temporary failures when a limit is exceeded. Issuing temporary failures is preferred because a legitimate sender will queue mail for later delivery. Queuing mail breaks the business model of a spammer. The spammer's goal is to deliver as many messages as possible in the shortest time possible, without investing in a messaging infrastructure – i.e., utilizing bot-nets. Queuing mail reduces throughput and requires more machines to send the same amount of mail.

The next level of sophistication with connection control is to apply time-based rules according to sender and recipient addresses as well as connecting IP address information. For example, by tracking the sending behavior of hosts that connect, spammers can be identified and blocked. For example, a sender who sends a significant number of messages whose recipients are invalid is probably either attempting to harvest the user list of the receiving organization by seeing which addresses are

rejected and which are accepted, or the sender is attempting to deliver unsolicited bulk email by guessing addresses. If the behavior is identified, subsequent attempts to send mail are met with temporary failures, which do not expose any future recipient addresses for harvest.

IP Reputation Services

IP reputation services are essentially sophisticated blacklists. A database of the reputation of sending hosts is maintained and queried to determine what type of a sender they are and what the risk level is of accepting connections from them. Mail from known spammers should be eliminated at connection time, because the likelihood is extremely high that the content from them is spam. The reputations of other senders may be combined with connection control to set a class of service for senders, thus preventing bulk senders, both legitimate, spammers and zombies, from monopolizing MTA resources. Commtouch has implemented an IP reputation service based on their Recurrent Pattern Detection (RPD™) system, which analyzes a large quantity of the entire network traffic on the Internet.

Sender Authentication

Sender authentication technologies when widely adopted defeat the ability to perpetrate fraud via email – that is, phishing. Sender authentication technologies allow the receiver of a message to determine that it is in fact coming from the domain (or sender) from which it purports to come. There are two categories of sender authentication technologies: path-based and cryptographic-based.

In path-based systems, such as Sender Policy Framework (SPF), Microsoft Caller ID, and Sender ID, special records in DNS are checked to determine whether the connecting host is allowed to send mail for that domain. The primary benefit of this system is that there is no cryptographic key management required to support it. The downsides are that it is not robust in the case of anonymous re-mailing like, such as mailing lists, and the message is not digitally signed.

In cryptographic-based systems, such as DomainKeys Identified Mail (DKIM), the messages are signed by the sender with a cryptographic signature that is verified with a public key distributed in a specially formatted TXT record in DNS. It does two things: it shows that the message is from the domain it claims to be from and it shows that the content of the message is unmodified. The cryptographic methods do not suffer from the problem of anonymous re-mailers. It does require deploying filters on both inbound MTAs for verification and outbound MTAs for signing.

“Sender authentication will almost certainly become a de facto standard part of the Internet’s email infrastructure over the next few years, but it will not stop the spam problem by itself.”

— ComputerWire/Datamonitor

Spam Scanning

This is the most computationally intensive method for fighting spam and there are a number of approaches and technologies which should be deployed.

To reduce the increase in false negatives, multiple spam scanning methods should be employed and coordinated to allow for easier as well as more elaborate and granular policy enforcement. For example, one community based (e.g., Cloudmark) and one network based (e.g., Commtouch, Eleven, etc) should be deployed together. Typically, the network based technologies are more effective at high volume spam campaigns and more effective against image-based spam because they rely less on the content of a message, than its distribution pattern. The community based technologies are better suited for identifying the lower-volume attacks like the so-called “Nigerian scam” messages that propose various business relationships or funds distributions from African leaders.

Sendmail’s products run anti-spam engines within the context of a policy engine. The policy engine handles the merging of results

among multiple engines. As long as the anti-spam engines run concurrently, the performance penalty is less than might be expected.

CONCLUSION

To effectively battle the onslaught of spam, businesses need to look at the problem holistically from both an architectural and technological view point. A properly architected gateway in combination with the right technologies, deployed in the right order, can significantly reduce the impact spam has on your business, users and system resources.

Gateway bounce management, content scanning and end user quarantines are all important best practices for blocking spam before it impacts your organization. Additionally, a defense-in-depth approach employing a number of technologies, including connection controls, IP reputation services and authentication in conjunction with the right combination of content scanning technologies, can drastically reduce the spam problem.

MESSAGING BEST PRACTICES EXPERTISE

Sendmail’s messaging experts can help you with your spam fighting strategy, best practices, solutions and implementation and support.

With 25 years leadership delivering innovative messaging technology, Sendmail ensures the protection and trust of employee and customer communications. Large enterprises in 33 countries, including most of the Fortune 1000, trust Sendmail to shield users from unwanted messages, defend the messaging infrastructure, stop data and privacy leaks, and effectively manage messaging to maintain brand and shareholder value, and support regulatory compliance.

SENDMAIL EMAIL SECURITY PRODUCTS AND SERVICES

To find out more about why businesses are turning to Sendmail to be their trusted messaging advisor, solution provider and implementation support partner, please call: Tel: +1-87-SEND-MAIL (877-363-6245) or +1-510-594-5400 (outside U.S.).

Sendmail, Inc.
6425 Christie Avenue,
Emeryville, CA 94608
Tel: +1 888 594 3150
Fax: +1 510 594 5429
www.sendmail.com