Cisco Email Security: Understanding Context Adaptive Scanning Engine (CASE)

Contents

Introduction

The increase in the volume of blended threats has been dramatic. Many of the most significant virus outbreaks in the past two years have been associated with spam delivery – meaning the virus payload creates an army of “zombie” computers – that are used to send spam, phishing, spyware, and even more viruses. Email-borne spyware has been doubling every six months, and it is not uncommon for spammed URLs to install “keyloggers” that steal usernames and passwords. Viruses can even be used to create a network of zombies to launch a massive distributed denial of service attack, such as when the Mydoom.B variant took SCO’s website offline with a coordinated assault.

What is driving the sudden increase in blended threats? In short, it’s the money. As first-generation anti-spam techniques (like blacklists and content filters) have been more widely deployed, traditional methods (like sending spam from a fixed bank of servers containing an “offer” in the text of the message) have become less profitable. With more networks using anti-spam technology, fewer “simple” spam messages make it past spam filters and into the recipient’s inbox. This hurts spammers’ profit margins and has forced them to adapt to these changes.

Spammers handled this situation in two distinct ways:

1. They are sending even more spam with the hope that what they lose in delivery rates, they will make up in volume.
2. They are turning to blended attacks to disguise their messages and increase their profit per message.

The second technique often becomes a criminal activity. Organized crime networks have been established to execute attacks and profit from viruses, phishing, and other threats. In 2004, an individual named John Dover was arrested after trading over two million credit card numbers, which were stolen through phishing attacks.

The techniques used in blended attacks have also become increasingly sophisticated. The Sober.N virus employed email, web downloads, trojans, and zombies. Traditional content analysis filters are no match for these intelligent threats. Many users of first-generation anti-spam filters

have found that they need to spend increasing hours “training” their filters or writing new rules. However, despite these efforts, their catch rate and throughput are both declining. The result is that costs escalate as more systems are required to keep up with the load, while more administration time is used to manage each system.

Cisco Email Security has addressed these threats with a unique blended threat defense technology known as the Context Adaptive Scanning Engine (CASE). Cisco Email Security’s CASE technology is used to stop both traditional spam and sophisticated zombie-based attacks. This same scanning technology is also used to prevent viruses and malware as much as 42 hours ahead of signature availability – with a single unified scan for efficiency.

Understanding CASE, Detecting Blended Threats In Context

First-generation filters were designed to look at the content of a message and make a determination. For example, if the word “free” appeared in a message more than twice, along with the word “herbal,” it was probably spam. This approach is relatively easy for spammers to defeat by using hidden characters or numbers instead of letters, such as “f0r y0u” in place of “for you.” Second-generation techniques, like Bayesian filters, attempted to address this limitation by learning to differentiate the characteristics of spam and legitimate email automatically. But these techniques proved too challenging to train, too late to react, and too slow to scan.

Given the advanced obfuscation techniques used with today’s spam, state of the art filters need to examine incoming mail in full context. CASE uses advanced machine learning techniques that emulate the logic used by a human that is evaluating the legitimacy of a message. A human reader, as well as Cisco Email Security’s CASE technology, asks four basic questions:

1. Who sent me the message?
2. Where do the links in the message take me?
3. How was the message constructed?
4. What does the message contain?

To follow is an examination of each logical area evaluated.

Who?

As stated earlier, first-generation spam filters relied primarily on keyword searches to identify spam. In 2003, Cisco (IronPort) revolutionized the email security industry by introducing the concept of reputation filtering. While content filtering asked the question, “What is in the message?”, reputation filtering asks the question, “Who sent the message?”. This simple but powerful concept broadened the context by which threats are evaluated. By 2005, nearly every major commercial security vendor had adopted some type of reputation system.

Determining reputation involves examining a broad set of data about the behavior of a given sender (a sender is defined as an IP address sending mail). Cisco considers over 120 different parameters, including email volume over time, the number of “spam traps” hit by this IP, country of origin, whether the host is compromised, and many more. Cisco has a team of statisticians that develop and maintain algorithms, which process this data to generate a reputation score. This reputation score is then made available to the receiving Cisco Email Security Appliance (ESA), which can then throttle a sender based on their trustworthiness. In short – the more “spammy” a sender appears, the slower it goes. Reputation filtering also addresses the problems associated with surging email volumes by either rejecting or throttling connections before the message is accepted, thus dramatically improving the performance and availability of the mail system. Cisco

ESA reputation filters stop more than 80 percent of incoming spam, approximately twice the catch rate of competing systems.

Where?

While the combination of email content analysis and reputation was state of the art in 2003, the sophistication of spammer's and virus writer's tactics continues to grow. In response, Cisco (IronPort) introduced the notion of Web reputation – a critical new vector to broaden the context in which a message is evaluated. Similar to the approach used in calculating an email's reputation, Cisco Web Reputation looks at more than 45 server related parameters to assess the reputation of any given URL. The parameters include the volume of HTTP requests to the URL over time, whether the URL is hosted on an IP address with a poor reputation score, whether this URL is associated with a known "zombie" or infected PC host, and the age of the domain used by the URL. As with email reputation, this Web reputation is measured using a granular score, which allows the system to deal with the ambiguities of sophisticated threats.

How?

Another novel approach to Cisco Email Security's contextual analysis is to examine the construction of a message. Legitimate mail clients, such as Microsoft Outlook, construct messages in unique ways – using MIME encoding, HTML, or other similar means. An examination of the construction of a message can reveal a great deal about its legitimacy. A most telling example of this occurs when a spam server tries to emulate a legitimate mail client's construction. This is difficult to do, and an imperfect emulation is a reliable indicator of an illegitimate message.

What?

A full contextual analysis needs to consider the content of a message, but, as noted earlier, content analysis alone is not a sufficient approach to identifying illegitimate mail. Cisco Email Security's CASE technology performs full content analysis, using state of the art machine learning techniques. These techniques examine the content of the message and score it in various categories – is it financial, pornographic, or does it contain content that is known to correlate with other spam? This content analysis is factored into CASE along with the other attributes – the Who, Where, How, and What – to evaluate the full context of the message.

CASE In Action

Because of the breadth of data analyzed by CASE, the technology is used in a variety of security applications – including IronPort Anti-Spam (IPAS), Graymail, and Virus Outbreak Filters (VOF). The example below highlights how CASE is used to stop spam. The message's content is nearly identical to the organization getting phished, so content analysis of the message would not identify any threats. To content-based filters, this message appears to be a legitimate communication. To determine whether or not this message is spam, filters that rely primarily on the "What" could easily be fooled into recognizing the message as legitimate. However, an analysis of the full context of the message paints a different picture.

- The IP address of the sending mail server is suspicious – it has had a sudden surge in volume, and the domain, in return, does not accept mail.
- The URL of the email points to a server that appears to be in a consumer broadband network.
- The URL advertised in the message is different from the "actual" URL that the user is

navigated to when clicking on the link.

When all three of these factors are considered in context, it becomes clear that this is not a legitimate message, but is, in fact, a spam attack.

Traditional "Content Filters"
What CONTENT FILTERS Find

What? Message content legitimate.

Verdict: UNKNOWN

Context Adaptive Scanning
What CASE Finds

What? Message content legitimate.
How? Message construction emulates Microsoft Outlook client.
Who?
1) A sudden surge in the volume of email being s
2) In return, the mail server does not accept mail.
3) Mail server located in Ukraine.

Where?
1) A mismatch between display & target URL web
domain registered a day ago.
2) Website hosted on consumer broadband netwo
3) "Whois" data shows the domain owner as a kn
spammer.

Verdict: BLOCK

When CASE is used in Virus Outbreak Filters, the same scoring and machine learning capabilities are applied – albeit to a separately tuned data set. Virus Outbreak Filters are a preventive anti-virus solution offered by Cisco and powered by CASE technology. The Outbreak Filters solution scans messages against both "real-time" Outbreak Rules (issued by Cisco Talos specific outbreaks) and "always-on" adaptive rules (that reside on CASE at all times), protecting users against outbreaks before they have had a chance to form fully. CASE enables Virus Outbreak Filters to detect and protect against virus outbreaks in several ways accurately. First, CASE can quickly scan messages based on parameters such as file extension of attachment, file size, filename, filename keywords, file magic (the actual extension of a file), and embedded URLs. Because CASE technology analyzes messages to this level of detail, Cisco Talos can issue extremely granular Outbreak Rules, that accurately protect against an outbreak with minimal false positives. CASE can dynamically receive updated Outbreak Rules, which ensures that it protects against the latest outbreaks.

In addition to the analysis of messages based on Outbreak Rules, CASE technology also scans messages based on Adaptive Rules. Adaptive Rules are finely tuned heuristics and algorithms that examine incoming messages for malformation and spoofing characteristics indicative of viruses. In addition to these parameters, Adaptive Rules score messages based on their SenderBase Virus Score (SBVS). SBVS is a score similar to a SenderBase Reputation Score (SBRS), but with a ranking based on the likelihood that the sending party is sending viral emails, rather than spam. A majority of viral email is sent by previously infected "zombie" machines, so identifying and scoring these sending parties is an essential factor in catching viruses.

Cisco Email Security's CASE technology enables Virus Outbreak Filters to stop virus outbreaks well before traditional anti-virus solutions because the CASE examines messages in multiple

ways. It has the ability to analyze numerous characteristics of message attachments, message content, and message construction, as well as the ability to analyze messages based on their sender reputation. And, because CASE also acts as the IronPort Anti-Spam and Reputation Filters engine, a message only needs to be scanned once for all of these applications.

High Performance, Low Cost

The logic behind CASE technology can be very sophisticated, and therefore very CPU intensive to process. To maximize efficiency, CASE uses a unique “early exit” technology. Early exit prioritizes the efficacy of the myriad rules processed by CASE. CASE technology runs the rules with the highest impact and lowest cost first. If a statistical verdict is reached (whether positive or negative), no additional rules are run, saving system resources. The elegance in this approach is having a good understanding of the efficacy of each rule. CASE automatically monitors and adapts the order of rule execution as efficacy changes.

The result of early exit is that CASE technology processes messages approximately 100 percent faster than a traditional rules-based filter. This has distinct advantages for large ISPs and enterprises. But it also has benefits for small and medium businesses. The efficiency of CASE, coupled with the effectiveness of Cisco Email Security’s AsyncOS operating system, means that ESAs with AsyncOS and CASE technology can be implemented on very low-cost hardware – driving down capital costs.

Another way CASE technology translates to low cost is by eliminating administrative overhead. CASE is tuned and updated automatically, thousands of times each day. Cisco Talos provides engineers who are trained, multilingual technicians, and statisticians. Cisco Talos analysts have special tools that highlight anomalies in mail flow detected in any Cisco Email Security customer’s network, or global email traffic patterns. Cisco Talos generates new rules that are automatically pushed to the system in real-time. Cisco Talos also maintains a massive corpus of “spam and ham,” which is used to train various rules used by CASE. The automatically updated CASE rules mean that administrators don’t have to be tuning and tweaking the filter or spending time wading through spam quarantines.

Summary

Spam, viruses, malware, spyware, denial of service attacks, and directory harvest attacks are all driven by the same underlying motive – profits. These profits are attained either through the sale or advertising of merchandise or theft of information. Profits from these sales are driving increasingly sophisticated attacks, developed by professional engineers. Advanced email security systems need to analyze a message in the broadest possible context to counter these threats. Cisco Email Security’s Context Adaptive Scanning Engine technology asks the four basic questions: Who, Where, What, and How – to weed out legitimate messages from blended threats.

- “Who” is the email reputation of the sender – who sent the message.
- “Where” is the reputation of the source hosting the website – analyzing where the link would take you.
- “What” is an analysis of the content of the message – what the message contains (first-generation systems often rely solely on the” What” type of analysis).
- Finally, “How” is an analysis of how the message is constructed.

This basic framework of analyzing Who, Where, What, and How works equally well for stopping spam as it does for preventing virus outbreaks, phishing attacks, email-borne spyware, or other

email threats. The data sets and analysis rule sets are tuned specifically for each threat. CASE technology allows the Cisco ESA to stop the broadest range of threats with the highest possible efficiency by processing these threats on a single high-performance engine.