IronPort's Multi-layer Spam Defense Architectural Overview

WHITE PAPER

TABLE OF CONTENTS

Executive Summary

Email threats have expanded from nuisance spam to sophisticated blended attacks. IronPort Anti-Spam eliminates the broadest range of known and emerging threats.

IronPort Anti-Spam™ combines best-of-breed conventional techniques with IronPort's breakthrough context-sensitive detection technology to revolutionize the fight against email threats. Today's spam attacks have become too sophisticated for earlier-generation spam systems. These systems share a common weakness – relying heavily on analyzing content that can easily be manipulated by spammers. State of the art anti-spam systems must go beyond content examination and analyze messages in the full context in which they are sent.

As spam continues to evolve, near real-time rules will need to remain a critical part of the anti-spam equation – in order to successfully eliminate spam and blended threats. With spam on the rise, this type of multi-layer defense is critical to protecting networks worldwide.

INTRODUCTION

The volume of spam has been steadily increasing every year since 2002. In addition to sheer volume, the sophistication of spammer tactics has also grown. This flood of illegitimate email is propelled by a powerful motive– profit. Spammers make money from selling a wide array of marginal products – ranging from herbal supplements, low interest mortgages, and ergonomic mice, to criminal activities such as credit card fraud, pornography and illegal pharmaceutical sales. The profits behind these endeavors are being plowed back into new technology and infrastructure for delivering spam.

When spam initially became a problem, corporations and networks began to deploy first generation spam filters. These filters primarily relied upon heuristic analysis – looking at the words in a message and using a weighting system to create a probability that the message was spam. As these solutions became more widespread, spammers began to develop new, more sophisticated, tactics to circumvent the filters. This spawned a cat and mouse game – in which spammers would develop a new tactic to get past filters, then anti-spam vendors would add a new technique to their "cocktail" to stop the spammers' tactic, then spammers would come out with a new tactic to get past filters, etc.

Recently, spam has been using increasingly sophisticated obfuscation techniques and mutating faster than ever. Most spam now includes blocks of text that contain words known to score as "not spam" – often technical terms or a passage from a text book. Other tricks involve using words with white on white text or replacing letters with numbers (e.g., L0ve). Spammers have also become increasingly clever in using URLs. Some spam contains minimal content but includes a URL with a call to action, while other spam attacks host their spam URLs on the same servers used by legitimate websites – using free Web hosting services, like Geocities.

These obfuscation techniques have effectively defeated most content based filters. While most vendors still claim to have spam capture rates in the high 90's, in reality, their capture rate may be in the 80's (or worse). At the same time, content based filters have the challenge of occasionally deleting legitimate mail that happens to contain words associated with spam creating a "false positive".

The table on page 3 highlights the evolution of spam filtering, along with the limitations of each of the approaches. The first three generations each overcame weaknesses of the prior generation, but all of these approaches suffer

from a common limitation. Each approach can be circumvented by spammers because it relies on something the spammers themselves have full control over – the content of the message. This is like building a house on top of a weak foundation.

GENERATION	LIMITATIONS	EXAMPLE
1. Hueristics	**Spoofable–** spammers change words so filters don't recognize spam but humans do. **False positives–** legitimate email often contains "spammy" words.	"C H E A P V.i.a.g.r.a"
2. Signatures	**Spoofable–** 'Hashbusters' fool bulk detection systems by making spam look dissimilar. **Reactive–** writing signatures first requires collecting spam samples.	"Cheap Viagra – dgjk#"
3. Adaptive	**Spoofable–** Defeated by inserting 'good' words that only machines see. **High Overhead–** Adaptive learning systems, like Bayesian, are hard to train/maintain.	"Cheap Viagra here: http://abc.com Cancer, office, Shakespeare...."
4. Context Adaptive	**Emerging–** Requires extensive vendor investment in tracking email and Web reputation.	

FROM CONTENT TO CONTEXT

Maintaining consistently high spam efficacy requires a new approach to the problem. This approach should leverage the latest in adaptive learning technology, but be based on a more holistic understanding of the context in which a message is sent. Importantly, this technology must incorporate information that the spammer cannot influence. This includes tracking the identity and reputation of the email sender and the website advertised in the message.

The spam filtering technology employed in the IronPort® email security appliances uses a highly advanced, multi-layer approach to evaluate a message. IronPort's anti-spam solution moves beyond traditional content based analysis by analyzing four broad areas:

1. Who (what do we know about the sender)
2. Where (if the message contains links, what do we know about where those links go)
3. What (what is the nature of the contents of the message)
4. How (how was the message technically constructed)

By examining a broad set of data, beyond the mere contents of a message, IronPort's anti-spam system yields robust, highly accurate results that require

no administrator intervention. This technology is currently deployed at the largest ISPs and enterprises in the world, protecting millions of end-user mailboxes.

The First Question – Who is sending me mail?

When a message arrives at an IronPort appliance, before any processing begins, the IronPort sytem evaluates the nature of the sender. This process is called Reputation Filtering – a technique pioneered by IronPort more than three years ago, and subsequently adopted by every leading anti-spam vendor. The concept behind Reputation Filtering is simple but powerful – analyzing the traffic patterns and network characteristics of a given sender to determine trustworthiness. The foundation of any reputation system lies in the quantity, quality, and breadth of data tracked.

Quantity – IronPort's SenderBase® is the world's first, largest and most accurate traffic monitoring network. SenderBase collects data on more than 120 parameters from over 100,000 different networks to characterize the behavior of a sender. This network includes eight of the ten largest ISPs in the world and a wide array of large and small enterprises, distributed globally. This powerful network gives SenderBase a view into an astounding 25 percent of the world's email traffic. SenderBase traffic represents a very statistically significant sample-size resulting in the extremely high accuracy of IronPort Reputation Filters.

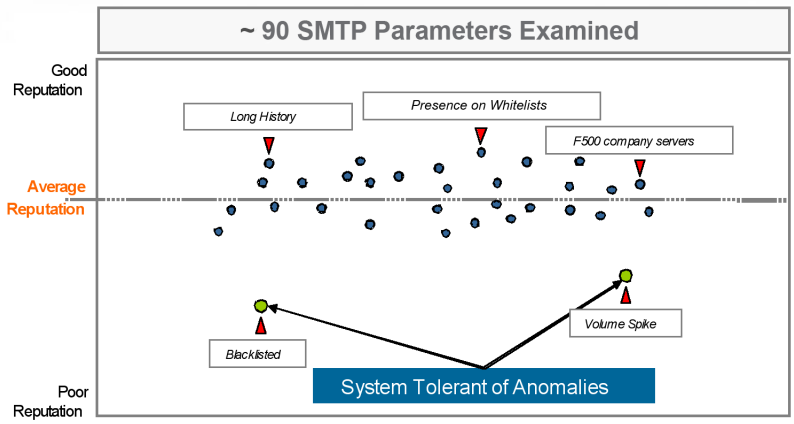
Quality – In addition to size and breadth of data, IronPort has developed a sophisticated data quality engine that allows SenderBase to account for data feeds from different sources with different circumstances – normalizing them for proper interpretation. (See the IronPort Anti-Spam Ecosystem section for more information).

Breadth – The data measured by SenderBase includes the global volume of mail being sent by a particular sender, how long the sender has been sending mail and at what volume, whether the sender accepts mail in return, what the country of origin is, and whether the sender's DNS is configured properly. These are all objective, network-based parameters that can be accurately measured.

Because they look at such a broad set of sender data, IronPort Reputation Filters are robust enough to overcome occasional outlying data points. This effect is illustrated in Figure 1.

Figure 1: ***Global Efficacy for Broad Threats***

Broad data analysis drives accuracy



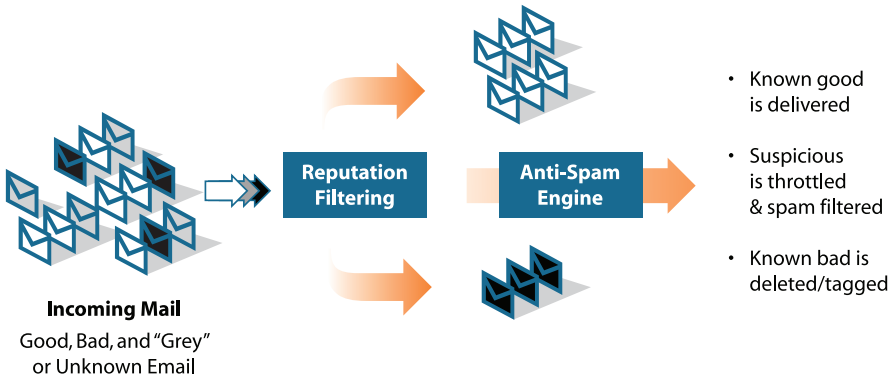
As an example, a sender who has a long history of sending reasonable mail volumes, accepts mail in return, and is a Global 2000 company – but happens to have a mis-configured DNS record – will still have a positive reputation, despite one or two questionable parameters. However, the sender who is sending 10 million messages per day, has just begun mailing on their IP, does not accept mail in return, is sending from a "zombie" PC, and is located in the Ukraine is likely to have a negative reputation. Consequently, mail from this sender can be stopped before it even enters the network.

DNS blacklists and whitelists were the predecessors to reputation systems and some reputation systems today are still based solely on this early generation technology. The advantage of a true reputation system is not only the breadth of data, but also granularity. Traditional blacklists are binary – a sender is either guilty (which means they are blocked) or not guilty (which means they can send as much mail as they would like). IronPort Reputation Filters offer higher granularity, measuring sender reputation on a scale of -10 to +10. This allows the IronPort appliance to deal more gracefully with ambiguity. IronPort Reputation Filters are linked to IronPort's unique rate limiting capability. This allows the IronPort appliance to intelligently "push back" or slow down a sender that appears suspicious but has not yet earned a reputation worthy of blocking. In short, the more "spammy" a sender appears, the slower they go. Having the ability to dynamically apply limits to new or suspicious senders allows the IronPort to greatly reduce the amount of incoming spam, without incurring false positives – because suspicious or ambiguous mail is slowed but not blocked.

In production for more than three years, IronPort Reputation Filters are so accurate they can stop 80 percent of incoming spam at the connection level. This powerful outer layer reduces email bandwidth consumption and

hardware administration costs by as much as 80 percent. This outer layer also serves as a valuable "shock absorber" – by dramatically reducing the number of messages that need to be scanned during a denial of service attack, spam outbreak or virus outbreak. This filtering is illustrated in Figure 2.

Figure 2: IronPort Reputation Filters are the outer layer of defense, stopping 80 percent of hostile mail at the door.



- IronPort uses identity & reputation to apply policy
- Sophisticated response to sophisticated threats

But the concept of reputation does not stop at the perimeter. The reputation of the sender is passed to IronPort's Context Adaptive Scanning Engine™, known as CASE. This the engine that looks at a broad set of data – including sender reputation – to evaluate a message in context and make a final determination of "spaminess".

The Second Question – Where do the links in this message take me?

In order to make money, spammers need a "call to action" in their message. This call to action may be a phone number to call to buy a product, a physical address to send money to or the ticker symbol of a penny stock the spammer wants you to buy. More often than not though, the call to action in an email message is a URL linking to a website with a product offer or malicious content. Over 85 percent of spam today contains a URL in the message.

Just like blacklists and whitelists of IP addresses three years ago, vendors are trying to address this problem by constructing blacklists and whitelists of URLs. This approach is like a "whack a mole" game, however, as spammers generate hundreds or thousands of URLs, often only for a few hours. By the time traditional URL blacklists list a new URL, the attacker has defrauded his victims and moved on to using a new URL. Similar to email reputation, solving this problem requires the ability to track the reputation of both the URL and the entity that controls it in near real time.

IronPort collects more than 40 different parameters to determine the reputation of a website. For example: How long has the domain been registered? Are the domain's whois records valid? What country is the website hosted in? What is the reputation of the network hosting the website? What is the global volume of requests to this site? How has that volume changed over time? What is the nature of the content on the site? What is the reputation of the mail server that sends URLs linking to the site?

This data is collected in SenderBase from more than 100,000 different networks in a similar manner as email traffic data. IronPort's statisticians have developed Web reputation algorithms, that are similar to the email reputation algorithms. These algorithms produce a Web reputation score, which is made available to the IronPort CASE for spam filtering.

The Third Question – How was this message constructed?

Today it is increasingly easy to buy an off the shelf "spamware" package to-generate millions of email messages. These packages are extremely powerful, but often leave traces that indicate the program generating the message. Spammers also have a vested interest in masking their real identites – and exploit the weaknesses of SMTP by forging elements of their messages.

Structural rules examine how the message is constructed, looking for subtle patterns that differentiate good mail from bad. For example, does the message contain signs of obfuscation – like legitimate text that is hidden using a font color that is nearly identical to the background color? Do the message headers contain the fingerprint of a known "spamware" toolkit – like Sendsafe used by spammers to send their messages? Structural rules also help identify signs of forgery. For example, does a message claim to come from a trusted webmail provider, but really originate from an entirely separate source?

The Fourth Question – What does the message contain?

While inadequate in and of itself, content analysis is useful when applied in the full context in which the message was received.

IronPort Anti-Spam includes advanced lexical analysis that examines the contents of each message and considers this in the context of who is sending the message, how it was sent and where links in the message point to. A message may contain the word "Viagra", but if it is coming from a source that is a known pharmaceuticals company, the positive sender reputation score will offset any suspicions raised by the content and the message will pass through. Similarly, a message that contains many financial terms

such as "mortgage" and "interest rate", but appears to be coming from a consumer broadband network that does not accept mail in return, will have a high likelihood of being spam. IronPort Anti-Spam has the ability to interpret all major international languages, including double-byte characters used in most Asian languages.

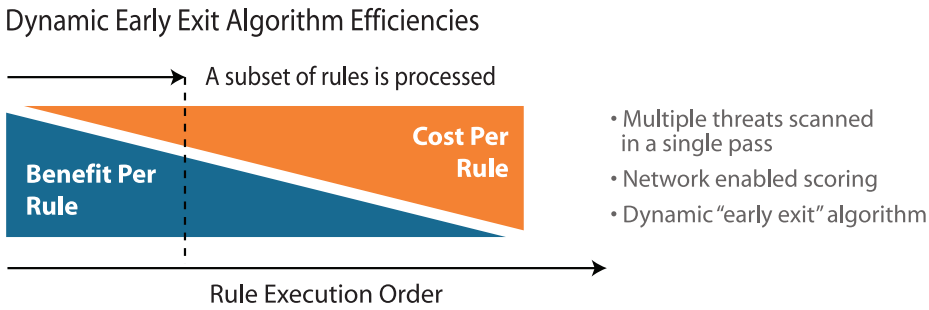
ANALYZING IN CONTEXT WITH CASE

IronPort's Context Adaptive Scanning Engine (CASE) pulls all of these questions together. By examining a message in its full context, considering who it is from, where the links point to, how it was constructed and what language it contains, the CASE is an extremely powerful machine learning system that makes accurate spam/not spam decisions. By examining a message broadly in its full context, it begins to emulate the logic that a human would use when evaluating an unknown message. Who is it from? Where does it take me? Does it look real? What language does it contain? As illustrated earlier in Figure 1, this broad contextual analysis allows the CASE technolgy to look beyond a few attributes, that might appear anomalous, and accurately classify messages as spam or not.

One of the challenges associated with contextual analysis is that a comprehensive examination of every message can be extremely computationally intensive. IronPort offsets this challenge by eliminating unwanted email as soon as enough information about a spam message is known to block it. Reputation Filters ensure that the CASE only examines the small percentage of mail that is not clearly known good or known bad. CASE technology uses a unique "early exit algorithm" to efficiently reach verdicts.

Early exit allows IronPort's CASE to stop scanning a message once a verdict is reached. By running the most applicable rules first, the majority of spam can be stopped without running the entire rule set. Two unique aspects of the CASE early exit system are: the order in which rules are processed is updated dynamically, and the early exit algorithm is applied to legitimate email as well as spam. This unique approach yields a massive increase in throughput for the system, allowing the CASE to process more than three times the throughput of traditional rules-based spam filters. The early exit concept is illustrated in Figure 3.

Figure 3: ***IronPort Anti-Spam Advantage: Performance***
Early Exit Accelerates Scan Time



THE IRONPORT ANTI-SPAM ECOSYSTEM

The tactics of spam are always changing, meaning a world-class anti-spam system must constantly be measuring and responding to these changing tactics, and have facilities to provide real-time updates to stay ahead of the flood. IronPort has developed unique technology and invested in large scale infrastructure to measure and characterize spam behavior, providing a dynamic stream of updates to its appliances in the field.

A critical component of anti-spam efficacy is the quality of the rules run by CASE. IronPort's Threat Operations Center (TOC) has built a very sophisticated system to measure and manage rule efficacy and to generate a constant flow of new rules to respond to the shifting tactics of spammers.

At the heart of the TOC, is a massive and highly diverse database. Data streams into the TOC from more than 100,000 different networks around the world, including very large entities such as eight of the ten largest ISPs in the world. This data feed includes SMTP and HTTP traffic data, utilized in the email and Web reputation systems, and also a stream of millions of spam messages from a variety of sources.

To be able to account for the varying quality and sources of this huge feed of incoming spam, IronPort has developed a data quality engine. This engine uses statistical techniques to compare the results of a given data feed with the characteristics of a known sample and then normalize the results. For example, a feed from trained and proven reliable human reporters may indicate a 90 percent probability of spam, but a feed from a large consumer ISP might only be a 50 percent probability of spam. If the same message shows up in both feeds the probability might grow to 96 percent probability of spam. The technicians and statisticians in IronPort's TOC have had years of experience properly interpreting and weighting different data sources.

This network of over 100,000 sources feed the world's largest "corpus" of email. The corpus contains messages from around the world that have been classified with certainty as either spam or legitimate email. The corpus is constantly updated with millions of new messages daily, automatically classified, and then verified with human oversight from the TOC analysts. The corpus is used to generate new rules automatically as well as manually.

The TOC contains rule writing technicians tasked with detecting the small subset of spam that automated systems fail to detect. These technicians are equipped with tools to group messages that share similar underlying characteristics using a patent pending technique called Feature Similarity Vectoring (FSV). Unlike "fuzzy checksum" approaches that rely on several message attributes to determine message similarity, FSV determines message relatedness by analyzing thousands of message attributes. By associating seemingly disparate messages, analysts are able to quickly write rules, based on the underlying attributes common across the attack.

Once a technician creates a new rule, it gets added to the body of rules processed and goes through a battery of tests to ensure that is accurate. Using advanced statistical techniques, the entire body of rules is repeatedly run against the corpus and each rule is assigned an optimal weight. Rules that are less effective are expired or dropped from the rule set. Rules are automatically ordered, based on their contribution towards catching spam. This dynamic ordering of rules is a key enabler for IronPort's unique early exit algorithm, described earlier.

The extensive technology and infrastructure of the Threat Operations Center creates over one hundred thousand new rules every day. These rules are sent to IronPort appliances using both "push" and "pull" updates. In some cases the appliances will pull new rules or launch a query to SenderBase about a particular sender. In other cases new rules are "pushed" to the appliance. The update mechanisms include HTTP and DNS text records. Rules on the system are automatically updated, deployed, cached, and expired – depending on the class of rule. This highly robust update schema allows the IronPort appliances to provide reliable protection, even if some or all of the centralized rule infrastructure ever became unavailable.

ENTERPRISE MANAGEMENT

When serving large enterprises, having cutting edge technology is obviously an important, part of a successful solution. But equally important is enterprise reporting and management tools to minimize administrator burden and help address the business case for the investment required in the

email security infrastructure. IronPort has developed a world-class reporting system that allows IT staff to measure the return on their investment, as well as advanced management tools to adapt to the varied needs of a global enterprise end-user population.

Building a truly enterprise-class reporting and management system is non-trivial. Enterprise reporting and management have been designed in to the IronPort appliance since their inception. Every IronPort appliance contains a real-time reporting system called Mail Flow Monitor™. Mail Flow Monitor gives a real-time view into what is happening in the system. Is mail backing up in a queue? Is the system being attacked by a spam outbreak or DDoS attack? The system automatically highlights anomalies and generates SNMP and/or email alerts as required. Mail Flow Monitor also provides a historical summary of how much traffic has been received, what percentage is spam, virus, blocked by reputation filtering, etc. These historical reports can be automatically generated and distributed periodically.

In addition to the on-box reporting of Mail Flow Monitor, IronPort offers powerful centralized reporting with Mail Flow Central™. Mail Flow Central pulls log data off of multiple appliances and stores it in a SQL database. This database can be queried using IronPort's simple Web based tools to generate historical reports, perform capacity planning, and support ROI analysis. In addition, Mail Flow Central has powerful message tracking capability that allows IT staff to easily see what happened to any given message. Tracking can be done by sender, by recipient, domain, size – virtually any message attribute. This unique capability reduces the trouble shooting burden significantly.

In addition to real-time and centralized reporting systems, IronPort has developed a family of end-user facing controls. IronPort Anti-Spam supports a simple outlook plug-in that allows end-users to identify and report missed spam at the click of a button. This spam is automatically routed back to IronPort and the filter algorithms are tuned based on the feedback.

The IronPort appliances also support a fully integrated end-user quarantine to store either "suspect spam" or "suspect and definite spam". Many customers who use the quarantine only do so for suspected spam and drop known spam because of the extremely low false positive rate of IronPort Anti-Spam. The quarantine can automatically generate a summary email that is sent to end-users with subject lines of all quarantined messages. If a user sees a subject of interest they simply click on the link and launch a familiar webmail interface. There they can view messages and release or delete them. Released messages are routed through the IronPort appliance, so it can automatically adjust its algorithms. Quarantine size limits can be set and messages are automatically purged. The quarantine application is fully integrated into the appliance.

SUMMARY

Today's spam attacks have become too sophisticated for earlier-generation spam systems. These systems share a common weakness – relying heavily on analyzing content that can easily be manipulated by spammer. State of the art anti-spam systems must go beyond content analysis and analyze messages in the full context in which they are sent. Maintaining leading efficacy also requires publishing high-quality rules in near real time. Rule quality is driven by the size, breadth, and quality of the data that feeds the rule generation system. Finally, the most effective rule development systems have humans in the loop – analyzing and responding to the last few percent of spam messages that escaped automated defenses.

IronPort Anti-Spam is unique in the industry – it analyzes messages in their full context, allowing the system to be very robust and accurate. IronPort has pioneered the concept of reputation filtering, starting with email reputation and more recently Web reputation. These two factors are very powerful components of full context analysis, because they are based on factors not easily controlled by spammers. IronPort has also innovated with its Context Adaptive Scanning Engine that examines email reputation, Web reputation, message construction and content as efficiently and accurately as possible. This system is supported by the industry's most sophisticated Threat Operations Center, which captures and processes massive quantities of data – to keep IronPort Anti-Spam one step ahead of ever-changing email threats.

IRONPORT®

IronPort Systems
950 Elm Avenue, San Bruno, California 94066
TEL 650.989.6500 FAX 650.989.6543
EMAIL info@ironport.com WEB www.ironport.com

IronPort Systems, a Cisco business unit, is a leading provider of anti-spam, anti-virus and anti-spyware appliances for organizations ranging from small businesses to the Global 2000. IronPort appliances utilize SenderBase, the world's largest email and Web threat detection network and database. IronPort products are innovative and easy-to-use—providing breakthrough performance and playing a mission-critical role in a company's network infrastructure.

IronPort is now part of Cisco.

CISCO