

The Challenges for Anti-Spam Technologies

Spam is a global problem that continues to grow at an alarming rate. It knows no national boundaries, comes in every language and every message format. The level of spam is growing every day and with it the definition of spam is also rapidly changing. Up until 2003, spam was defined as unsolicited bulk commercial email. Today, the definitions cover a vast gray area, which under the broad provisions of the CAN-SPAM Act of 2003, may even include legitimate solicited bulk mail.

Furthermore, what is spam to one person may be desired or even required reading by another. At the end of the day, an intellectual decision by the end-user must be made to complete the efficacy of any anti-spam solution.

Spam is costing enterprises and economies billions of dollars every year. The more than 15 billion spam messages sent every day cost European businesses an estimated \$2.5 billion and US corporations \$10 billion in 2003.

By January 2004, enterprise users were already reporting dramatic increases over previous years, complaining that more than 50% of their average daily incoming mail was spam. Others have suggested the figure is closer to 80%.

The key to saving valuable enterprise resources is to block spam and bulk messages before they are delivered to the recipient. With this dramatic increase, more and more companies are developing technologies to combat the proliferation of spam. Many of these solutions remain focused on spam determination based purely on the lexical analysis of the content of the message. Spammers meanwhile study these detection methods and continue to find ways to circumvent content-filtering based solutions.

This White Paper discusses the challenges of spam as faced by vendors, currently available detection technologies, and the effectiveness of current spam filtering solutions for enterprises and their recipients.

The main challenge faced by Anti-spam technologies is to develop an effective spam filtering system that must be able to see, in real-time, patterns in suspected spam messages to determine if an attack is underway. They must not be deceived by slight changes in the message. Furthermore, the technology must be scalable to handle both increased spam as well as the slower but more natural growth of an enterprise.

Increased Spammer Sophistication

Spammers continually find new "tricks" to make their email messages appear innocent so they can pass through spam detection applications. Their increasingly sophisticated methods for delivering spam include: sending the entire spam message as an image, slightly altering the content, subject line, or From field with spam message, including the use of non-English phrases or content, avoiding the use of known spam

"keywords," disguising the actual URL(s), using quotations from classic sources (e. g. novels and poetry) to intrigue the recipient and more.

Challenge: An effective spam filtering technology must be able to adapt to new spammer tricks and be proactive, rather than reactive, in the battle.

Minimizing False Positives while Maximizing Spam Detection

As a strategy to deliver high rates of detection, many anti-spam technologies are tuned to aggressively analyze keywords and phrases. The result is often an unacceptably high level of false positives. This happens when a solution focuses mainly on detecting spam based on keywords. If these spam "keywords" are present, the message is considered spam, even if in its context the message is actually a legitimate business communication.

Most anti-spam applications, with varying degrees of sophistication, falsely assume a message is spam based mainly on a lexical analysis of the content, without balancing the context for the particular enterprise or industry in which the message was sent. The more these anti-spam applications try to find spam in every message based solely on the aggressive analysis of keywords and phrases, the more false positives they report.

To avoid cases of false positives, administrators may configure the anti-spam solution to analyze incoming messages less aggressively. While this does successfully lower the instances of false positives, it has the opposite result of raising the level of false negatives whereby more spam infiltrates the enterprise and challenges the very reasons why the enterprise turned to the anti-spam solution in the first place.

Challenge: A successful anti-spam solution must completely eliminate the synchronization between high detection rates and high false positives, delivering extremely effective detection with close to zero false positives.

Reducing IT Overhead

Enterprise-defined policies seeking to outwit spam quickly become obsolete as spammers change the content of the message and cleverly disguise the source. Therefore, IT managers using most anti-spam applications are required to constantly define and redefine enterprise policy and create new rules in an attempt to combat spam. Many solutions are largely ineffective until they are trained, and then still require excessive IT resources to keep spam definitions current. Solutions that are difficult to deploy and require manual updates similarly tax IT resources at an unacceptable level.

Challenge: An effective anti-spam technology must be easily deployed, automatically updated and require little or no intervention from the IT department. The system should require little or no training period during which the enterprise continues to be inundated with spam. While the solution should be a learning system, it should offer its own effective

solution that can be fine-tuned by additional IT input, but is very effective even without enterprise-side intervention.

Customizing Spam Filtering Management

One person's spam is another person's favorite newsletter. Business-critical communication for one user will be considered spam by another. Solutions that rely on a lexical analysis of the contents of a message often apply the same value to all users, regardless of their personal interests or preferences. An oncologist will not consider a message about the latest developments in breast cancer research as spam, while the purchasing department will welcome receiving information about the availability of the latest ink refills.

Challenge: Any effective anti-spam solution must be able to treat copies of the same email message sent to various users differently; so that in one case, the message is correctly sent to the recipient's inbox, while in the second case, the exact message is correctly deleted.

Overview of Key Anti-Spam Technologies

Following is a partial list of key technologies currently being used to protect organizations and individuals from receiving spam messages in their Inbox folders.

- Blacklist/Whitelist
- Content Filtering
- Heuristic Detection
- URL Detection
- Bayesian Filters
- Challenge/Response
- Hybrid Detection

Blacklist/Whitelist

Commercial blacklists offer enterprises a list of domains that are extensively known as spammers or as open relays for spammers. In principle, this may appear to be a valuable resource. However, this service is often arbitrary, is not based on any industry-wide standard for what should or should not be included on this list, relies on opinion rather than statistical analysis and has been shown to result in high levels of false positives.

Enterprise whitelists, specific to each organization, help implement enterprise-side policies to fight spam. Here too, while this would appear to be a valuable resource in combating incoming spam it represents a drain on IT resources constantly required to define and redefine filters to approve trusted sources. Furthermore, if the enterprise restricts incoming mail to those on the whitelist, it risks missing important business-critical messages.

At the same time, if it uses only the whitelist to permit non-spam messages while using other spam detection systems for checking non-trusted sources of email, the whitelisting efforts are not actually effective in combating spam. The enterprise may have succeeded in permitting known messages to enter the system, but has done little to effectively analyze and detect messages that are not from trusted sources. Also, the issue of customizing the definition of spam is not addressed because the enterprise may either whitelist a source for all users or not whitelist the source for anybody. Once on the list, the message will arrive in the Inbox folders of all recipients, even those who may not want it.

Content Filtering

Theoretically, a technology that scans the contents of an email message might appear to be an adequate solution to spam. However, solutions relying only on content-based detection face several difficulties in determining when a message is spam. For example, content filtering solutions cannot process images. Therefore, any spam containing an image is permitted through because the content filtering solution cannot "read" the image.

Similarly, if the message is not in English, the filter cannot classify the message as spam. To counter this, the filter will have to possess several "dictionaries" of foreign languages in order to be able to "read" the message; a cumbersome and expensive proposition.

Another problem with content filtering solutions is that they offer one definition of spam for all users. A message considered spam for one user is considered spam for all. Sophisticated spammers can easily outsmart content filtering solutions by not using known "key words." Because these solutions seek to scan the content, but not the *context* of the message, it results in high levels of false positives. The only way to lower the high level of false positives is to lower the aggressive classification criteria. While this does lower the level of false positives, it also results in a lower rate of spam detection. In short, content filtering offers a stagnant, limited classification capacity.

Heuristic Detection

Recognizing that content filtering solutions cannot be maintained manually, Heuristic detection solutions were developed to enable automatic, dynamic content filtering by assigning weight, or values, to keywords. Unlike the "plain vanilla" content filtering, the Heuristic detection solution is a learning system. It can be updated, with new keywords added automatically. In theory, this solution also sounds like an effective way to combat spam. But it too suffers from many of the limitations found in content filtering solutions. Heuristic detection cannot scan images, is language dependent (cannot detect non-English spam) and the rate of false positives remains relatively high.

URL Detection

URL detection is considered to be on the cutting-edge of spam detection. Since many incoming spam messages include a link that the spammer hopes the recipient will click, URL detection seems to be a reliable way to detect the domain names of spammers. Indeed, while this solution is adequate, it does have its own limitations. URL detection solutions cannot detect URLs contained in images and, perhaps even more critical, it can do nothing for emails that do not have any links coded into them (the email offering millions from a Nigerian bank account in exchange for the recipient's bank details being just one infamous example).

Like other solutions that do not fully appreciate the sophistication and determination of spammers, URL detection cannot anticipate tricks designed to evade this form of detection. As spammers learn to evade this method, for example by putting a URL in an image or disguise the source URL via chain of proxies, spam detection vendors relying solely on this technology find it impossible to keep track of the latest URL-related tricks.

Bayesian Filters

Bayesian filters offer yet another version of spam detection, taking content filtering to an even higher level. In theory, this might be considered a very strong identification technique, as it is customizable, learns as it goes along and makes the job of the spammer more difficult. However, Bayesian filters, like all spam solutions that rely on word analysis as a base, cannot scan images and is often language-dependent (interpreting only English messages).

Another problem is that the further the process of spam definition is removed from the recipient, the less accurate it is. If the IT department is determining the spam definition for the enterprise, individual users are likely to be less satisfied by the level of spam protection they are receiving. If the ISP or vendor is making the definitions, the accuracy of the spam definitions suffers even more. Finally, regardless of who makes the definitions, Bayesian filter-based solutions require extensive interaction and customization.

Challenge/Response

Challenge/Response solutions are designed to avoid spam from "hit-and-run" spammers who send massive amounts of messages with bogus email addresses, by challenging the senders to respond to a validating message initiated by the anti-spam solution before forwarding non-trusted messages to the recipients. While this method has proved somewhat effective against spammers who have "fled the scene" after sending bulk messages, it does great injustice to legitimate senders...exactly those mailers who do not deserve any punishment.

Similarly, newsletters sent in bulk and customarily sent from an automated address and provide a different address for inquiries. When the Challenge/Response solution requests confirmation from this legitimate newsletter, it will not receive a response and will therefore reject this wanted correspondence as spam.

Hybrid Detection

Finally, another trend in the industry is to use Hybrid detection methods. This method uses, in various combinations, several of the above techniques. The theory being that what might not be caught by one method will be by another detection engine. Again, while this seems to be logical, studies have shown that the results, if anything, are much less accurate than many of the above techniques used individually. Hybrid detection technologies tend to produce confused and contradictory information that leaves the IT department unsure of what to trust. One method may determine something is not spam and send it to the second technology. This secondary technology may then determine that the message is spam, thus assuming that the first technology diagnosed a false negative.

In fact, by combining solutions and definitions, the overall level of false positives is increased, with little additional spam protection.

Spam Patterns

Each spam outbreak contains at least one repetitive value, a pattern that will repeat itself in every message of the outbreak. This is true even though spammers have already learned how to mutate and alter the contents of almost every email in a massive spam attack, thus making them appear innocent enough to pass through content filtering detection solutions.

Simple examples of spam patterns could be header information (i.e., From address, Domain name, Sender IP, Subject string); Envelop information (i.e., Mail-From, RCPT-To) or the source URL. More sophisticated patterns would typically include sub-values of strings, headers, envelop or the message-body, the coupling of values, placement, etc.

The challenges facing pattern detection techniques include the constantly changing nature of messages; that each outbreak contains new and unique patterns and that there can be no dependency on knowledge from previous outbreaks.

Pattern detection requires a high degree of sophistication, in order to deal with matching across three main categories:

- Identical match
- Approximate match
- Cross match

Furthermore, substantial scalability and performance issues must be taken into account.

There are nearly 1 million new patterns every day and the database must contain and continually update many millions of patterns while unused old

patterns are removed. Database redundancy is therefore a vital prerequisite.

The Commtouch Approach

Commtouch has developed a unique and highly successful response to all these challenges with its Recurrent Pattern Detection (RPD™) technology, which focuses on detecting patterns in spam attacks, rather than on a lexical analysis of the contents of individual email messages. It is content-agnostic and can detect spam in any language, format or encoding method.

It is a modular integrated solution, designed to meet enterprise-specific spam detection and protection requirements while meeting the needs of both the small business sector and large multi-national corporations, ISPs and OEMs.

The Commtouch Anti-Spam solution consists of three major components:

RPD technology is responsible for proactively probing the Internet to gather information about massive spam outbreaks from the time they are launched. This patent-pending technology is used to identify recurrent patterns that characterize massive spam outbreaks. On average, the RPD technology can recognize unique recurrent patterns in new spam attacks within the first 1.5 minutes. Because it does not rely on a lexical analysis of the contents of the message, the RPD technology can detect spam in any language and in any message format (image, text, HTML). Additionally, this technology is equally effective for single and double byte encoding.

The Commtouch Anti-Spam Detection Center hosts and employs the RPD technology to detect, identify, analyze and classify unique recurrent patterns of spam. It holds a vast database of already classified patterns, and delivers up-to-the minute spam-detection services to Commtouch customers. In January 2004 the Commtouch Detection Center identified a daily average of nearly 1 million new and unique spam patterns, representing a more than 50% growth over June 2003. Classifications of new spam patterns are added every day to the millions of patterns already identified on previous days.

The Commtouch Spam Detection Engine, an OEM spam detection module, enables application and appliance vendors to enhance their products by incorporating the spam-blocking capabilities of the Commtouch Detection Center. The Spam Detection Engine's main function is to receive email messages from the integrated application and return the spam-related status of these messages.

It does so by creating digital keys representing specific message characteristics and polling the Commtouch Detection Center with that data. The Center returns a spam-related status report to the Detection Engine indicating the spam classification status of the various digital keys within the Center's database.

Spammers cannot outwit the Commtouch RPD technology because it is not based on a lexical analysis of the message-content and because the RPD technology was designed to constantly identify new patterns that were not

previously categorized. Meaning, the spam-patterns of the Commtouch database is dynamically updated in real-time by a fully automated and highly scalable system.

Key Benefits of the Commtouch RPD Technology:

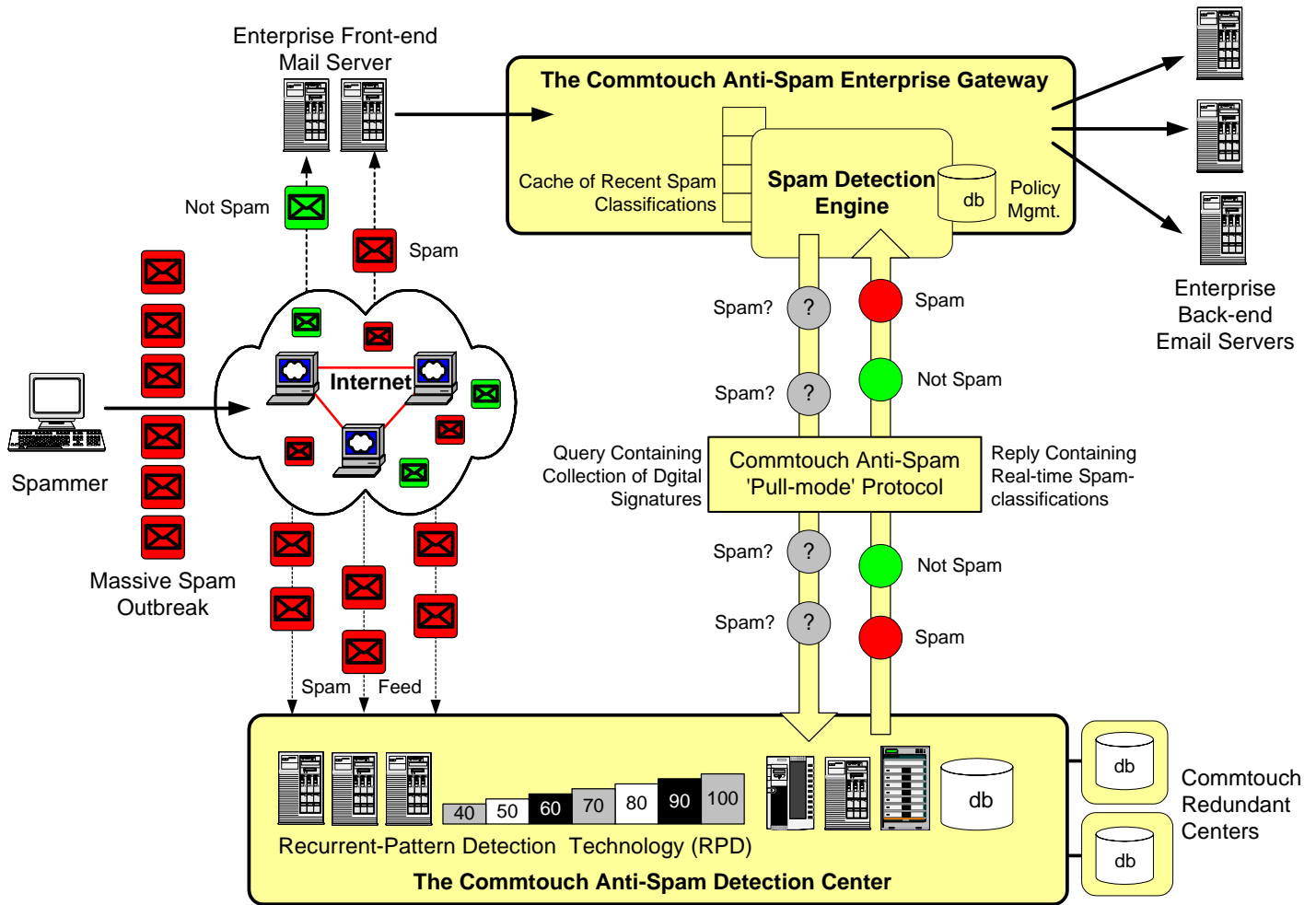
- Provides real-time protection against spam
- Offers a high spam-detection rate
- Reduces False Positives to almost zero
- Detects spam in every language and message-format
- Adapts to spam evolution through the lifecycle of a massive spam attack

How Commtouch Protects Customers World-wide from Spam

(Refer to diagram)

1. A spammer sends out a massive spam attack over the Internet. Within minutes, the Commtouch Anti-Spam Detection Center has already received and classified the recurrent patterns of the spam attack.
2. Meanwhile, new incoming messages arrive at the enterprise mail relay or front-end mail server that passes traffic to the Commtouch Anti-Spam Enterprise Gateway. After it is checked against enterprise policy, user rules, or the local spam detection, if still considered non-trusted, a query containing a collection of digital signatures representing only characteristics of the message, is sent to the Detection Center.
3. Within a few milliseconds, the Detection Center classifies the message and sends a reply to the Commtouch Detection Engine and Anti-Spam Application (the total round-trip time is 300 ms, excluding Internet latency).
4. The Commtouch Anti-Spam Enterprise Gateway applies predefined blocking policies (if spam: delete, quarantine, or send to user's Junk Mail folder; if not spam: to user's Inbox folder).
5. The Gateway stores the information in a local detection cache, making future local classification even more efficient.

Building Blocks of the Commtouch Anti-Spam Solution Using RPD™ Technology



Summary

Most anti-spam solutions rely on technologies that continue to prove ineffective against the growing onslaught of massive spam attacks being launched on a daily basis. By relying on a lexical analysis of the content of a message, most anti-spam solutions are unable to detect spam hidden in images or in non-English correspondence. Similarly, by aggressively defining the keywords regularly used in spam, these same solutions generate unacceptably high rates of false positives to achieve a high spam detection rate.

To counter this spam epidemic and its impact on enterprise users, Commtouch has developed the Commtouch Anti-Spam Detection Solution, utilizing the advanced spam detection and classification capabilities of the patent-pending RPD technology, which identifies recurrent spam patterns of massive spam attacks within 1.5 seconds of the outbreak.

RPD technology offers:

- Proactive Spam Detection
- Multi-language spam detection
- Multi-format message detection
- High detection rate (over 97%)
- Low level of false positives

Because the Commtouch Anti-Spam solution does not rely on a lexical analysis of the content of a message, the linkage between high spam detection rates and false positives does not exist. The Commtouch solution delivers extremely high detection rates with almost no false positives.

Commtouch enables ISPs to protect their subscribers by matching values of characteristics taken from new incoming messages to subscribers with classified patterns in the Detection Center and to distinguish between "good" and "bad" email messages. Similarly, Commtouch enables enterprises to protect their recipients by applying its own spam detection classifications at the local and remote level.

The Spam Detection Engine consists of a set of APIs, describing how to receive characteristics of incoming messages as input from applications, and how to interpret the spam-related status of non-trusted messages as real-time output of the Commtouch spam-detection services. Depending on the existing architecture of the OEM vendor's desired approach, the Detection Engine may be configured to receive deterministic-only results that classify messages as either confirmed-spam or non-spam, or to receive deterministic as well as statistical results, categorizing messages based on the likelihood that a message is spam. The Detection Engine module is easy to integrate, requiring only two DLL files and a User's API Guide. No additional resources need be invested by Commtouch OEM partners or their customers.



Recurrent-Pattern Detection (RPD™) Technology

About Commtouch

Commtouch® Software Ltd. has been a vendor of email technology, messaging applications and comprehensive messaging platforms to large ISP and Telco organizations since 1991. Publicly traded since 1999 (NASDAQ: CTCH), Commtouch delivered a highly scalable global messaging service, managing over 30 million hosted mailboxes, and adding as many as 140,000 unique mailboxes a day.

As a dedicated email service provider to tens of millions of users, Commtouch found it necessary to design an effective strategy to eliminate millions of unsolicited commercial email messages daily. The Company's core competencies - derived from vast hosted messaging experience as a global ASP and developer of email messaging applications - enable the company to deliver a unique anti-spam solution that outmaneuvers and overpowers today's sophisticated spam attacks. A combination of proprietary and patent-pending technologies enable Commtouch to accurately detect, classify, and block spam attacks as soon as they are distributed over the Internet.

For additional information contact your local Commtouch Sales representative or check our web site at www.commtouch.com.

© 1991 - 2004 Commtouch Software Ltd. All rights reserved.

The Commtouch Anti-Spam Enterprise Gateway is a licensed product featuring proprietary, patent-pending technology. All information contained in this document is protected by international copyright treaties. Commtouch Anti-Spam and RPD are trademarks of Commtouch Software Ltd. Microsoft, Microsoft Outlook, Microsoft Exchange, Active Directory, and Microsoft MSDE are trademarks and/or registered trademarks of Microsoft Corp. All other trademarks and registered trademarks are the property of their respective owners.