WHITE PAPER: Commtouch RPD™ Technology
Network Based Protection Against Email-Borne Threats

Spam, phishing and email-borne malware such as viruses and worms are most often released in vast quantities in a relatively short period of time, causing global damage of tens of billions of dollars annually. Although many methods to combat these threats have been developed and tested throughout the past several years, a common deficiency of most of these methods is that they lack the ability to adapt quickly enough to the rapid change of distribution and infiltration techniques invented by spammers and virus authors.

Commtouch's Recurrent Patterns Detection (RPD) technology, based on the identification and classification of message patterns delivers the highest threat detection capabilities. The objective of this document is to discuss the characteristics of these threats, the challenges facing technologies that aim to mitigate these often malicious attacks, and describe how the Commtouch solution protects against all types of email-borne threats.

The Commtouch approach is based on the understanding that all threat outbreaks share some common characteristics, including:

- Most email messages within the outbreak have been altered to make it difficult to set blocking rules based on lexical analysis.
- Most outbreaks include millions of email messages to maximize the highest possible response rate and the greatest ROI for the attacker.
- Most outbreaks are released within a short period of time, requiring a real-time solution to detect the outbreak to limit or avoid the damage that can be incurred.
- The originators of the attacks invest heavily in disguising their origin to make it difficult to track the message back to them.

 In more detail, the following challenges were identified:

Spam Detection

Spammers typically produce spam for commercial gain rather than malicious intent. Although the purpose may be to sell a product or service to recipients, spammers often use unethical methods to acquire email addresses of recipients and to distribute mass mailings. In composing spam messages, spammers use sophisticated tactics to evade existing spam detection applications. This includes covering the tracks to the spammers, manipulating or hiding the commercial URLs, use of non-English words and phrases and a host of other methods.

Typically a massive spam outbreak will only last a few hours and be launched from a network of 'zombie' machines. To complicate the detection process, each message within the massive spam outbreak can be composed differently and employ more than one evasion technique.

Phishing Detection

Password Harvesting, or phishing messages, are typically sent for the single purpose of identity theft. There is an underlying intent within each message to violate the privacy of the recipient and commit fraud. The goal of a phishing message is to fool the user to go to a web site which will request them to enter private information. By using highly effective social engineering methods, these messages target users, often with a sense of urgency, to believe they have arrived at a legitimate site such as their bank or recognized online vendor to enter their private information. The sender is then able to gather personal information such as credit card numbers, passwords, social security or identification numbers and other private information.

Phishing messages appear to be from genuine or credible sources. Like spam, phishing messages can be sent in any language or format in attacks that typically last only a few hours and are usually launched from an army of 'zombie' machines on the Internet.

Virus Outbreak Detection

Typically, email-borne virus or worm outbreaks are created and released for the malicious purpose of either committing fraud, industrial espionage, or in the hope of gaining financial benefits by gathering private information that can be sold to spammers or those with criminal intent. Like spam and phishing messages, each virus message can be packed differently in terms of its content and the characteristics of the executable files that include the virus. However, email-borne viruses and in particular worms, can be received from legitimate and trusted email sources that might have been previously infected and were unintentionally distributing the virus to others.

Like spam and phishing, email-borne virus attacks often last for very short durations. In the case of viruses, users are exposed and unprotected during the first hours of the attack because most anti-virus defenses depend heavily on the use of a database of signatures that identify the threat by matching it with already-known characteristics. Recently, virus writers have become even more sophisticated and have begun distributing multiple modified instances of the same virus within a series of outbreaks to evade heuristic systems and to maximize the impact before new signatures are propagated.

Message Patterns

Massive outbreaks which distribute spam, phishing, and email-borne viruses or worms, consist of many millions of messages intentionally composed differently in order to evade commonly-used filters. Nonetheless, all messages within the same outbreak share at least one and often more than one unique, identifiable value which can be used to distinguish the outbreak.

For example, in the case of spam the objective is to lead the recipient to the same commercial web sites that can be classified as spam. In doing so, different spam attacks are often launched from the same network of zombie machines that can be blacklisted. In the case of phishing, recipients are lured to voluntarily disclose personal and confidential information via clever social engineering methods and the objective is often to lead the victims to same faked URLs. Email-borne viruses always contain the same malicious code (otherwise it is a different virus or another instance of the same virus). All these are

recurring values of typical outbreaks. These values are called the 'message patterns' of the outbreak. Any message containing one or more of these unique patterns can be assumed with a great deal of certainty to be part of the same outbreak.

Message patterns are extracted from the message envelope, headers, and body with no reference to the lexical meaning of the content. Thus pattern analysis can be used to identify outbreaks in any language, message format, and encoding type. Message patterns can be divided into distribution patterns, which determine if the message is 'good' or 'bad' by analyzing the way it is distributed to the recipients, and structure patterns, which determine the volume of the distribution.

The challenges of message pattern classification determining which message patterns identify outbreaks without generating cases of false positives, and how to extract and analyze these patterns before the outbreak wanes. Most outbreaks have a relatively short lifecycle measured in only a few hours. Therefore, any solution that does not detect and classify messages in real-time will only be effective towards the end of the outbreak, when most of the damage has already been done. All outbreaks attempt to disguise messages as legitimate email correspondence pretending to arrive from trusted sources and therefore, solutions that are based on pattern analysis must be able to tell the difference between 'good' and 'bad' patterns and avoid making mistakes.

The challenges are made more complex by the fact that each new outbreak usually introduces completely new patterns that were not previously analyzed and are therefore unknown to the pattern analyzer. Pattern detection represents a new and greater understanding of how email-borne threats are created and propagated. Because tactics for distributing spam, phishing, and email-borne viruses and worms are constantly evolving, it is necessary to proactively identify new and unique patterns in real-time in order to determine new outbreaks as they are released to the Internet and begin targeting recipients.
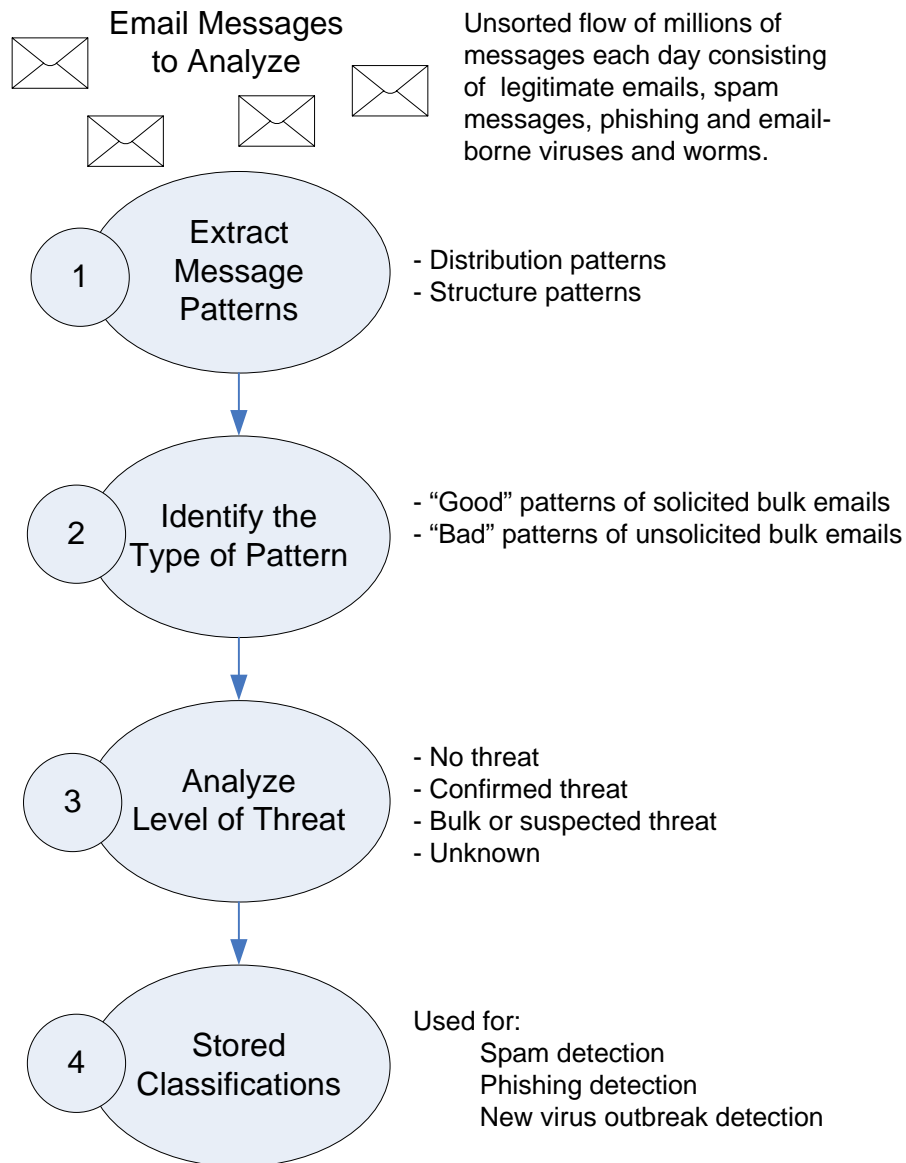
Recurrent-Pattern Detection (RPD™) Technology

Recurrent Pattern Detection (RPD) technology detects and classifies all types of email-borne threat patterns in real-time. RPD is hosted by the Commtouch® Detection Center, which proactively analyzes vast amounts of Internet traffic in real-time.

RPD, a patent-pending technology based on Commtouch's U.S. patent #6,330,590, extracts and then analyzes relevant message patterns, which are used to identify massive email-borne outbreaks. RPD classifies both distribution patterns and structure patterns and the analysis results are stored in a vast warehouse of classifications. In addition to identifying new threat patterns, RPD is also used to modify or enhance classifications of already-identified message patterns.

RPD is designed to distinguish between the distribution patterns of solicited bulk emails which represent legitimate business correspondence, from those of unsolicited bulk emails by applying a reverse analysis. The results of this analysis are 'bleached' message patterns belonging to 'good' messages such as popular newsletters, mailing lists, etc.

Commtouch uses the RPD technology in a highly scalable environment to deliver extremely high performance rates by analyzing many millions of new patterns each day, (24x7x365). New outbreaks are identified within minutes since they are launched on the Internet globally.

The RPD technology was designed to be fully automated and requires no human intervention. To ensure maximum privacy and business confidentiality, RPD was designed to analyze hashed values of message patterns and not the 'open' values nor the message content.

Email Messages to Analyze

Unsorted flow of millions of messages each day consisting of legitimate emails, spam messages, phishing and email-borne viruses and worms.

1 — Extract Message Patterns
- Distribution patterns
- Structure patterns

2 — Identify the Type of Pattern
- "Good" patterns of solicited bulk emails
- "Bad" patterns of unsolicited bulk emails

3 — Analyze Level of Threat
- No threat
- Confirmed threat
- Bulk or suspected threat
- Unknown

4 — Stored Classifications

Used for:
    Spam detection
    Phishing detection
    New virus outbreak detection

RPD identifies nearly 100% of incoming threat messages with almost no cases of false positives. It is language-agnostic and is equally effective for all message formats and encoding types.

Summary

To effectively combat email-borne threats, a successful solution must address a growing number of challenges. Commtouch's RPD is a proactive detection technology that continues to outwit those who continue to invent new methods to propagate email-borne threats because it does not rely on the contents of the email and therefore, it is able to detect spam in any language and in every message format (including images, HTML, etc.), non-English characters, single and double byte, etc.

RPD is also unaffected by the lack of signatures for new email-borne viruses and worms and is capable of detecting these malwares within minutes of the attack's launch.

RPD technology offers:

- High spam detection rate with almost no cases of false positives
- Early detection of virus threats
- Protection against phishing attempts
- Content-agnostic threat protection
- Multi-language threat detection
- Multi-format threat detection

For these reasons, the RPD technology is the best companion to multi-prone messaging and security applications. RPD is available for fast integration in various ways, the most common of which is to embed the cross-platform shared object managed with a simple-to-use API.

Because RPD uses pattern analysis, it provides the best protection of investment for enterprises and vendors of messaging and security applications.

About Commtouch

Commtouch Software Ltd. (NASDAQ: CTCH) is dedicated to protecting and preserving the integrity of the world's most important communications tool -- e-mail. Commtouch has over 15 years of experience developing messaging software and is a global developer and provider of proprietary anti-spam and Zero-Hour virus protection solutions. Using core technologies including RPD (Recurrent Pattern Detection™), the Commtouch Detection Center analyzes billions of email messages per month to identify new spam and malware outbreaks within minutes of their introduction into the Internet. Integrated by more than 40 OEM partners, Commtouch technology protects thousands of organizations, with over 50 million users in over 100 countries. Commtouch is headquartered in Netanya, Israel, and has a subsidiary in Mountain View, Calif. For more information, see: **www.commtouch.com**. The site includes the Commtouch online lab detailing spam statistics and charts.

Trademarks and Licensing Agreement

© 1991 - 2005 Commtouch Software Ltd. All rights reserved.
Protected by U.S. patent #6,330,590.

The Commtouch Anti-Spam Enterprise Gateway is a licensed product featuring proprietary, patented technology. Commtouch Anti-Spam is a trademark of Commtouch Software Ltd. For more information, visit our website: http://www.commtouch.com/

All information contained in this document is protected by international copyright treaties. No information may be copied or reproduced without the express written consent of Commtouch Software Ltd. Any duplication, transmission by any method, or storage in an information retrieval system of any part of this publication for other purposes other than those stated above is strictly prohibited without the specific written permission of Commtouch Software Ltd. This includes, but is not limited to, transcription into any form of computer system for audio, text, print, or visual retrieval. All rights under federal copyright laws and international laws will be strictly enforced.

Microsoft, Microsoft Outlook, Microsoft Exchange, Active Directory, Microsoft SQL Server, and Microsoft MSDE are trademarks and/or registered trademarks of Microsoft Corp. Lotus, Lotus Notes and Domino are registered trademarks of International Business Machines Corporation in the United States, other countries, or both. All other trademarks and registered trademarks are the property of their respective owners.