Email Security Solution - Overview

As the damage caused by email-borne threats such as spam, phishing and malware (including viruses, worms, etc.) continues to grow in magnitude as well as in sophistication, effective solutions need to deliver high detection rates with minimal mistakes while providing an adequate response to the following core requirements:

- **Real-time detection and blocking** because most attacks last only a few hours and involve the release of tens of millions of emails containing malicious content or spam. If the solution does not work in real-time to detect and block outbreaks proactively, it will simply be reacting towards the end of the attack, when most of the damage has already been done.
- **Language and content-agnostic** because damaging threats are often found in emails that do not contain English characters or messages that use images rather than text. Solutions that focus on the content of a message are likely to miss these threats, decreasing their overall effectiveness and increasing instances of false negatives received by recipients.
- **Adaptive to new methods and tricks** because malicious hackers and sophisticated spammers change their tactics of distribution and infiltration constantly in order to fool and evade current technologies.

The purpose of this document is to examine the current challenges imposed on email security applications and to outline one aspect of Commtouch approach's to spam and malware detection. Commtouch has created two engines that work together to deliver the highest possible detection rates. The RPD engine, detailed in this document, focuses on analyzing message patterns rather than on a lexical analysis of the content of messages. The LocalView engine, which enhances Commtouch's patented Recurrent Pattern Detection (RPD) technology by incorporating a rule-based engine. The LocalView engine is documented in the *LocalView Rules Management Manual*.

Key Email Security Challenges

The Commtouch solution focuses on providing detection for three types of email-borne threats: spam, phishing, and viruses.

- **Spam:** When composing spam messages, spammers use sophisticated tactics to evade existing spam detection applications. This includes masking the originator or sending machine of the spammers, manipulating or hiding 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.
- **Password Harvesting, or phishing messages**, are typically sent for the single purpose of identity theft. Some regard phishing messages as a subset of spam while others focus on the malicious aspect and refer to them as yet another type of malware. Phishing authors employ highly effective social engineering methods that are almost impossible to resist on the recipient-end. Like spam, phishing messages can be sent in any language or format, are distributed in attacks that typically last only a few hours and are commonly launched from an army of 'zombie' machines.

- **Email-borne virus or worm outbreaks** are created and released for malicious purposes. Like spam and phishing messages, each virus message can be different in terms of its content and the characteristics of the executable files that contain 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 are 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 instances of the same virus within the same outbreak 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 the 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 include 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.

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 recurrent message patterns in outbreaks, rather than on a lexical analysis of the contents of individual email messages. The aim of RPD is to identify and classify all types of email-borne threats and outbreaks. It is content-agnostic and can therefore detect spam and phishing in any language, format, or encoding method, while identifying patterns of new email-borne viruses and worms for which signatures have not been made.

RPD, a patent-pending technology based on Commtouch'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 reconfirm and enhance the classification of already-identified message patterns.

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). On average, new outbreaks are identified within minutes from the time 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.

RPD technology is hosted at the Commtouch Datacenter to proactively analyze vast amounts of Internet traffic in real-time to classify message patterns. The Commtouch Datacenter also hosts the Commtouch Classification Warehouse, a vast repository of threat patterns. Commtouch has developed two products for integrating RPD into your messaging system:

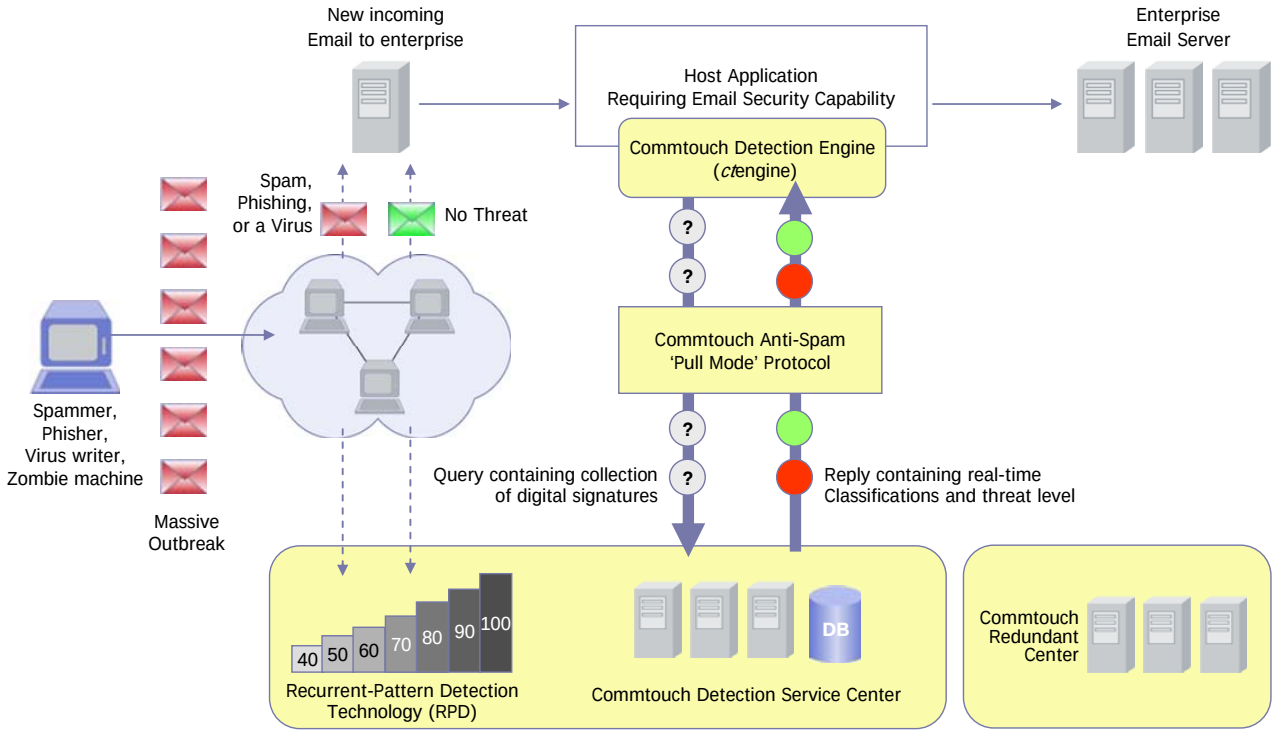
- ctengine: an email-borne threat detection engine, which is a cross-platform software module that can easily be integrated into third-party applications using a simple-to-implement API.
- ctasd: a plug-n-play, email-borne spam and malware outbreak detection daemon that combines your current core messaging network infrastructure with advanced detection and classification capabilities.

Both products have a very small footprint and can be integrated into a third-party host applications quickly. Contact your Commtouch Sales Representatives for information on how to deploy ctengine or ctasd.

When integrated with the host application, Commtouch's products analyze email messages, extract message patterns, and formulate queries to the remote Commtouch Datacenter over a proprietary protocol. It then returns accurate classifications for spam, phishing and virus threat levels to the host application to apply an action (such as release

to the recipients, delete the message, quarantine for second opinion or delayed decision, etc.).

Typically, ctengine partners include vendors and integrators of security appliances, anti-virus programs, firewalls, routers, modems, mail servers, security gateways, desktop applications, etc. who wish to offer email-borne threat detection services in their product or to enhance their already-existing anti-spam, anti-virus, or anti-phishing capabilities. The following diagram illustrates in a nutshell the Commtouch threat detection and blocking solution:



1. A new massive attack is launched over the Internet. Within minutes, RPD has identified the threat pattern and forwarded it to the Commtouch Datacenter, which classifies and stores the recurrent message patterns of the attack in a vast repository of message patterns.
2. New incoming messages arriving at the enterprise mail relay or front-end mail server are passed to ctengine/ctasd for analysis.
3. Message patterns are then checked against enterprise policy and user rules (optional).
4. A query containing a collection of digital signatures representing only patterns of the message is sent to the Commtouch Datacenter.
5. Within a few milliseconds, the Commtouch Datacenter classifies the message patterns and sends a reply to ctengine or ctasd, which then sends the classification to the application hosting it. The size of the query is about 500 bytes and the total round-trip time is ~300 ms, excluding Internet latency.
6. The host application applies predefined blocking policies (delete, quarantine, or send to user's personal quarantine or Inbox folder, etc.).

7. ctengine/ctasd stores the information in a local detection cache, making future local classification even more efficient and especially effective if the communication with the Commtouch Datacenter is temporarily unavailable while new messages continue to flow into the organization from other routes.

The Commtouch Datacenter is highly scalable, includes provisioning for redundancy and load balancing, and is highly secured from external attacks. Because of the robust way in which the Center was deployed, since the first day of operation, the Datacenter has never experienced a single down time period. New patterns are added to the Commtouch Datacenter's vast message pattern repository in real-time and made available immediately to all ctengine units worldwide.

Conclusion

The Commtouch approach to threat detection and protection is based on an understanding of the fundamental challenges constantly posed by today's malicious hackers and sophisticated spammers. More importantly, the Commtouch solution is prepared for future tactics and methods these groups will develop in the future and because of its modular nature, is easily adapted to future requirements and developments in the industry.

Commtouch responded to the realization that most threat outbreaks cause the most damage in a relatively short period of time from release, by developing a real-time detection solution. Typically, within the first minutes of the release, Commtouch has already proactively identified the outbreak and is able to classify and instruct the host application to block any messages from the outbreak before they reach recipients.

Because the Commtouch approach does not focus on contents analysis, it is completely irrelevant whether malicious hackers vary the contents of the message, use non-English characters, images, single or double byte encoding, etc. Even when virus authors release multiple instances of the same virus within an attack, Commtouch is able to track and block all variations.

About Commtouch

Commtouch® (NASDAQ: CTCH) provides proven Internet security technology to more than 150 security companies and service providers including 1&1, Check Point, F-Secure, Google, Microsoft, Panda Security, Rackspace, US Internet, WatchGuard and Webroot, for integration into their solutions. Commtouch's GlobalView™ and patented Recurrent Pattern Detection™ (RPD™) technologies are founded on a unique cloud-based approach, and protect effectively in all languages and formats. Commtouch's Command Antivirus utilizes a multi-layered approach to provide award winning malware detection and industry-leading performance.

Commtouch technology automatically analyzes billions of Internet transactions in real-time in its global data centers to identify new threats as they are initiated, enabling our partners to protect end-users from spam and malware, and ensure safe, compliant browsing. The company's expertise in building efficient, massive-scale security services has resulted in mitigating Internet threats for thousands of organizations and hundreds of millions of users in 190 countries.

Commtouch was founded in 1991, is headquartered in Netanya, Israel, and has a subsidiary with offices in Sunnyvale, California and Palm Beach Gardens, Florida.

Commtouch Inc.
292 Gibraltar Drive,
Suite 107,
Sunnyvale, CA 94089, USA
Tel: +1 650.864.2000
Fax: +1 650.864.2002

Commtouch Software Ltd.
4A Hatzoran St
P.O. Box 8511
Netanya 42504, Israel
Tel: +972-9-863-6888
Fax: +972-9-863-6863

Trademarks and Licensing Agreement

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

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.