



XML MALWARE:

CONTROLLING THE PROPAGATION OF MALICIOUS SOFTWARE WITHIN SERVICE ORIENTED ARCHITECTURES

A WHITE PAPER

Forum Systems, Inc.

BOSTON, MA
95 Sawyer Road, suite 110
Waltham, MA 02453

SALT LAKE CITY, UT
45 West 10000 South, suite 415
Sandy, UT 84070

TOLL FREE
1-866-333-0210

WWW.FORUMSYSTEMS.COM



TABLE OF CONTENTS

XML: A CARRIER OF MALWARE	3
A DECEPTIVE PATH TO DETECTION	3
<i>Data Interception</i>	4
<i>Decoding Binaries</i>	4
<i>Data Format</i>	5
<i>Content and Transport Encryption</i>	5
<i>Policy Enforcement and Administration</i>	5
CONCLUSION	5

Author: Walid Negm
Vice President of Marketing
Forum Systems Inc.

Release Date: 6/01/2005





XML: A CARRIER OF MALWARE

A virus can seek novel ways of spreading to hosts by adapting to its environment and mutating over time. In cyber space, a software virus will also discover ways of spreading, deliberately as well as accidentally. New “infection vectors” may result from additional pathways whereby malicious code can travel or more deliberately when the exploit capability of a virus is enhanced.

Most recently the former has been in the news. Instant Messengers, smart phones and Bluetooth-enabled gadgets have allowed malicious code to propagate and spread. Seamless connectivity between devices, servers and applications make it easier for us to communicate and get things done. On the flip side malicious code writers don’t have to adapt their exploits, they rely on old tricks and newly sprouted pathways.

Extensible Mark-up Language (XML) technology is accelerating the degree of enterprise interconnectivity and integration. By describing information in such a way that everything can mutually agree on meaning and context this technology is a primary enabler of Web and desktop applications as well as sophisticated electronic supply chains. XML is becoming pervasive.

XML was initially conceived to describe textual information. A purchase order, credit application, business contract etc. When we talk about textual information, we mean data which contains primarily printable characters. However, our world is filled with non-textual information:

<ul style="list-style-type: none"> • <i>WEB PAGES</i> • <i>IMAGES</i> 	<ul style="list-style-type: none"> • <i>AUDIO/VOICE/MUSIC</i> • <i>VIDEO</i> 	<ul style="list-style-type: none"> • <i>ANIMATION/MULTIMEDIA</i> • <i>2D/3D DATA / VIRTUAL REALITY</i> 	<ul style="list-style-type: none"> • <i>APPLICATION DATA</i> • <i>PROGRAMS/SCRIPTS</i>
---	--	--	--

Text-only XML is insufficient for robust business transaction and communication. Insurance claims require accident photographs, online auctions are enhanced with pictures of the items being traded, medical records may need x-ray scans and Computer Aided Design graphics are core to collaboration. By creating a seamless fabric of interconnectivity and interoperability, XML offers malicious users better odds when exploiting software. An XML document can be disguised as a denial of service attack, a springboard for malicious software or siphon of private information.

It’s no wonder that hackers have already recognized XML as a path of least resistance to spread malware. Managing an exploit does not require a lot of imagination. A software virus can be embedded within an XML document, making it unrecognizable to antivirus scanners. In fact, as more documents are stored in the XML format, the likely hood of a virus traveling completely unseen across a local or wide area network is already happening.

A DECEPTIVE PATH TO DETECTION

The propagation of malicious software through XML and Web services allows executable programs to perform unauthorized and destructive acts. The standard approach to block malicious programs is to automatically filter data prior to reaching its final destination. Commercial virus scanners detect malicious code circulating across the Internet using a database of well-known Spyware, Viruses, Worms and Trojan code signatures. The scanner systemically compares byte-sequences of suspect executables (e.g. scripts, macros and software programs) to their updatable signature database.





As XML and heavily networked service oriented architectures are deployed within the enterprise the risks associated with malicious software need to be re-evaluated. Organizations should install data filters across all known permutations of infection vector pathways, and in particular those routes that are taken by XML data and Web services.

The following section describes five areas that need be proactively addressed in order to control the propagation of malicious software within service oriented architectures:

DATA INTERCEPTION

Some malicious code scanners are smarter than others and have been packaged within applications such as email clients. Others are preemptive, sitting well ahead of the application at the edge of the network as gateways. Depending on the deployment model, arresting the propagation of active-content within XML documents can occur at a number of stages:

1. Transport: FTP, SMTP, HTTP, Message Queues etc.
2. Operating System: Windows, Linux, Mac OS etc.
3. Application Container: ActiveX, Microsoft Outlook, Web Browser, Business Applications etc.

Organizations need to put together a vulnerability containment strategy that supports integrating XML antivirus scanning at multiple stages to decrease the probability of infection.

DECODING BINARIES

In the world of binary data there are a number of methods to maintain the integrity of content across interactive applications, operating systems and networks. The first is to encode binary data using specialized algorithms (HEX, Base64, Base32 and ASCII85) making binary data look the same across platforms.

The second approach is to create a consistent way to package multi-media content with regular text. The Multimedia Internet Message Extensions (MIME) standard was developed to support the exchange of binary data such as graphics, audio and video files. The idea behind MIME is to package non-textual information along with text. MIME is just one of the ways that XML can package binary data. MIME also works for XML messages that have been wrapped in SOAP (Simple Object Access Protocol). Alternately, the World Wide Web Consortium (W3C) has proposals for binary encodings within SOAP payloads, referred to as SOAP Message Transmission Optimization Mechanism (MTOMS) and XML-binary Optimized Packaging (XOP). Microsoft also has a private specification known as Direct Internet Message Encapsulation (DIME).

Organizations should keep an eye out for the encoding algorithms and message formats that make sense for their business needs. One approach is to block non-MIME messages. This not only filters non-relevant binary data but also ensures interoperability between communicating parties. More importantly, malicious code in binary data will not be detected unless decoding has occurred and organizations should ensure antivirus scanners are updated with the latest specifications.





DATA FORMAT

Antivirus scanners go beyond looking at byte sequences by adding heuristics, anomaly detection and knowledge of certain content types. For example, Microsoft Office 2003 XML documents places macro-code at specific locations. It would be more efficient to target those locations for inspection rather doing a brute force examination of the entire document. This would reduce errors in detection and is important in high volume data centers. Organizations should ask themselves if their antivirus solution can parse XML 1.0 and SOAP 1.1/1.2.

CONTENT AND TRANSPORT ENCRYPTION

Confidentiality of Web services can be ensured in-transit using standards such as SSL/TLS (Secure Sockets Layer/Transport Layer Security) or more persistently using XML Encryption. In both cases, “termination” of the encryption algorithm has to be performed before any data examination can begin. XML Encryption is becoming a critical approach to protecting data without losing the confidence that it could be intercepted as it travels across multiple stakeholders. Organizations should mandate their antivirus scanner either support XML Encryption or integrate with an application that can pre-process encrypted XML.

POLICY ENFORCEMENT AND ADMINISTRATION

Blocking attachments and alerting users of harmful messages are some examples of email specific security policies. In the case of mission critical Web services and time-sensitive transactions, the enforcement and administration policies will be more sophisticated and tailored to the business process. Security filters should include rules to extract binaries, modify headers and quarantine content. Continuous monitoring of data flows such as origin, destination and other logging and auditing capabilities should also be part and parcel of XML antivirus capabilities.

CONCLUSION

Enterprises need to be one step ahead of the hacker and new security risks. XML, Web services and Service Oriented Architectures (SOAs) introduce an expanded threat profile which exposes data flows and new targets for exploitation. As the number of nodes, pathways and interaction patterns within service-oriented applications sky rocket, organizations must look at modern techniques to protect their data. An XML Firewall enforces deep content inspection, data-level access control and is able to control the propagation of malicious software within service oriented architectures.



ABOUT FORUM SYSTEMS

Forum Systems, Inc. is the Leader in Web Services Security™ with a comprehensive suite of trust management, threat protection and information assurance solutions for the automated Web. Forum flexible hardware, software and embedded products make vibrant business communications possible by actively protecting XML data and Web services across networks and business boundaries. Forum's products are used by Fortune 1000 companies and winners of Network Computing Magazine's Well-Connected 2004 Award, Product of the Year 2004 Award, Editor's Choice 2003 Award, Network Magazine's Product of the Year 2003 Award, DEMO 2004 Invitation and InfoWorld LEADERBOARD 2004.