

High-performance protection at the network edge—what, why and how

With more demands being put on lightweight network-edge hardware to provide security as well as connectivity, a new hybrid approach combining simplified malware detection with URI filtering can offer excellent proactive protection without overburdening the hardware or the administrator.

By Artur Kocharyan, Product Manager, Sophos

High-performance protection at the network edge—what, why and how

Summary

With more demands being put on lightweight network-edge hardware to provide security as well as connectivity, a new hybrid approach combining simplified malware detection with URI filtering can offer excellent proactive protection without overburdening the hardware or the administrator.

Combining solidity and efficiency—the protection conundrum

The malware epidemic continues to flourish around the world, with every step forward in technology and connectivity polluted and abused as soon as it becomes popular. As networks and working practices evolve to take advantage of the latest technologies, new protective strategies must evolve to provide the best possible level of security. Monitoring network gateways and links for malicious traffic has been a standard practice for some time. But as traffic demands and network complexity increase, simply proxying traffic through traditional security filters running on traditional processing hardware is becoming an ever more expensive option, less than ideal in many environments.

The ideal point to detect and remove threats is at the edge of the network itself, but the devices sitting there often have limited resources and are not ideally suited to conventional scanning techniques, based on local databases of specific and generic detection signatures. In modern multi-site businesses, branches are often connected via lightweight routers or UTMs with limited resources. Even in major sites deploying full-scale appliances performance is under pressure from increasing demands.

A new approach is required to provide fast, accurate threat blocking that will be effective on hardware with limited memory and processing power, and with little opportunity for local

updating. A lightweight, flexible but still powerful approach, combining high levels of protection with minimal resource usage, is the only viable solution to the problem of efficient, effective threat defense in an environment where resources are at a premium and throughput latency carries a heavy cost.

A web full of threats—the modern malware attack vector

In the early days of malware, the primary infection vector was the floppy disk; this was how data was transferred from one machine to another, and so malware followed the same route to find new victims. As email became ubiquitous and we began exchanging files as attachments rather than writing them to solid media, malware joined in and infected attachments became the transmission method of choice, with worms incorporating their own SMTP engines, developing social engineering techniques and spreading wildly. For some time now this vector has been in decline; although it continues to represent a threat, it is mainly limited to exploiting document formats, as savvy administrators have long since learned to simply block all executable attachments automatically. The main danger of email these days is in the links it may carry to malicious content, hosted out on the internet.

These days everything is on the internet. We not only store our own data online and exchange information with colleagues and clients over the web, we also require access to the mass of collective information posted online by others. Restricted web browsing is no longer an option in most businesses, as any user in any way hampered in accessing the information required to get his or her job done is an employee not working to maximum efficiency. It's no longer simply a question of monitoring the gateways at the edges of corporate networks either, as multi-

site businesses become the norm, and remote and travelling workers require equal access to both internal company networks and the web at large. The network boundary has become a porous, vaporous idea rather than a tangible line in the sand.

Malware has of course followed this trend, and the internet is steadily drowning in dangerous content. The proportion of infected sites, whether malicious by design or legitimate sites compromised to host threats, grows exponentially. Social engineering used to lure us to these attack points grows ever more sophisticated, taking advantage of every emerging trend in communication, from email to the deluge of social networking systems, each of which seems doomed to become swamped in spam links to risky sites from the moment its first early adopters begin to sing its praises.

As malware itself is ever more cunningly crafted to elude detection and sneak precious data from infected systems, so the lures and tricks used to bypass our human defenses grow more devious day by day. Every possible weakness of the human mind, from greed to fear to simple curiosity, has been targeted as a way of persuading us to follow a link, allow a script to run, enter a password and throw ourselves and our machines open to whatever the malware creators have in mind. Most worryingly for businesses, malware and the associated infection techniques have become increasingly targeted to penetrate specific companies, trick specific users and steal specific, often highly valuable corporate data. While endpoint-level protection on the desktop provides a strong defense, most businesses will want to provide maximum security, with minimal impact on worker productivity, by filtering the bulk of threats at their network boundaries and at strategic points between network nodes.

A wall full of gaps—the modern business network edge

In the early days of business networks, the

corporate LAN was like a castle, sealed off securely from the outside world with a single way in or out, which could be heavily protected and monitored. These days such an approach is impossible, with users needing to connect in and out of the network, and with each other inside it, in a vast number of ways. Internal mail systems and data storage must be accessible externally for home workers or those on the road, not just from traditional computers but from smart-phones, PDAs and an ever growing range of devices. Clients and partners need to access online resources not just to read pages of text, but to interactively, post data and content and pass all kinds of traffic on and off web-facing servers. Even within the corporate network, different nodes and sites must be interconnected smoothly and transparently, but still need to be protected from the outside and each other. In many situations, monolithic, expensive appliance-based solutions can no longer provide the required protection without seriously impeding the flexibility of the business.

In such situations, the deployment of multiple, inexpensive lightweight devices at network boundaries and between nodes seems the ideal solution. With built-in firewalling on routers commonplace and lightweight UTM devices becoming ever more complete packages, the inclusion of malware detection is an obvious and logical step up. These devices represent an ideal bottleneck at which threats can be filtered out of the traffic flow, but by design are unsuited to the traditional threat detection model.

Standard malware detection techniques require ample processing power and memory space, and also need regular updating to maintain complete efficacy against the latest emergent threats. Desktop-level protection provides a solid last line of defense, with minimal impact on users thanks to the combination of modern, efficient scanning and high-powered hardware. But this last line should be shielded by additional layers

of protection, to keep users from the impact of malware attacks and the intrusion of detection alerts, keeping their systems and their working time efficient and profitable as well as safe. Many other modern devices also function as full-spectrum computers with far fewer resources, notably mobile devices, and these too require protection that does not overburden their limited memory, storage and processing power, adding another layer of complexity to the design of full-canopy protection, and these devices would also benefit from a different approach to security. Combining full-strength endpoint protection with additional layers that complement each other, covering the gaps in the armor, can provide a solid barrier against attack. But achieving the right level of overlap shifts much of the filtering burden to devices at the network boundaries.

In most lightweight network devices, the kind of resources required are simply not available, and updates represent an even bigger hurdle. Updating firmware on a UTM or router is a labor-intensive, generally manual task with an element of associated risk; any such update can lead to problems and loss of precious connectivity. Such a task needs to be performed as seldom as possible, while conventional anti-malware technology requires updates ever more frequently. Although core engine improvements may not need to be applied as often as detection updates, which in many cases appear hourly or even more often, engines still need updating once a month or so to add new functionality and enable proper detection of the latest generations of threats. This sort of schedule is well outside the acceptable boundaries for updating firmware-based devices.

At first glance, combining the requirements of these devices—the need for rapid throughput of high levels of data on low-powered, seldom-updated hardware—with the needs and powers of conventional anti-malware solutions seems an impossible task.

Compromise and avoidance—how the UTM problem has been approached

A number of solutions have been proposed and implemented to provide anti-malware protection in lightweight network-edge devices, with varying degrees of success. The simplest and most obvious have also suffered the most predictable shortcomings. The initial response to the requirement for network edge scanning was to skirt the issue entirely by imposing a proxy, rerouting traffic off the lightweight device and through a conventional scanner system running on a full-powered secondary machine. This allows for the most comprehensive detection the provider of the detection technology can produce, with full access to the required resources including frequent updates, but imposes a heavy burden of cost and equipment as well as less than optimal latency compared to the throughput capabilities of the lightweight device.

Other implementations have provided a standard file scanning engine with a pared-down, “core” set of detection definitions, speeding up throughput to an acceptable level and minimizing the updating requirements. This provides a basic level of protection against the most common threats, but has some fairly obvious negative implications for completeness of detection. A slightly more successful alternative has been to develop a simplified engine capable of processing the data streams passing through lightweight devices at reasonable speeds with regular-expression-based detection methods targeting known threat types. Some variants of this approach have even been able to introduce limited updating onto the device. Even the most sophisticated systems using such techniques offer no more than a cut-down version of the protection of a full traditional signature scanner, with incomplete proactive protection against new-emerging threats, and still impose a burden on resources and require physical access for regular updates.

Another option is the use of online resources, with the scanning system referring to detection and

threat data stored “in the cloud”. With some cloud-based systems requiring considerable amounts of traffic to transfer detailed file characteristics and detection information back and forth, latency can become a serious problem, especially with large numbers of files passing through a device and requiring off-system lookups. A per-file lookup approach, which imposes heavy increases in traffic volume and slows the device throughput as well as potentially compromising the privacy of sensitive data, is no silver bullet.

The complete solution—a hybrid approach

Taking all these conflicting requirements and capabilities into account, a set of rules for the ideal approach begins to emerge:

- » For maximum economy and efficiency, malware filtering needs to take place on the lightweight device, ruling out the proxying method.
- » To maintain the required throughput level, the scanning engine must be lightweight and capable of processing large amounts of data with limited resources of memory or CPU.
- » To keep the memory footprint down, large amounts of detection data should not be stored on the device; while to avoid firmware updating requirements, improvements to the engine logic must be applied live and remotely, leaving the hard-coded components integrated into the device firmware untouched.
- » Finally, complex remote lookups of each and every file need to be avoided to minimize latency and added traffic, so a second form of threat detection is required—filtering not by file, but by URI.

With a large and sophisticated web monitoring system identifying and cataloging infected sites, a quality threat lab can build a comprehensive database of known bad URIs, rapidly spotting and responding to new outbreaks. This database can be checked at high speed from the detection system of a lightweight device as it processes the HTTP data passing through it, and malicious sites can be blocked in real time. Filtering the source

of malware rather than filtering by file recognition saves both time and effort; the complex obfuscation techniques employed by some malware seeders—such as server-side polymorphism, where files are tweaked and repacked frequently to minimize scanner detections—are completely bypassed. As the web is increasingly the vector of choice for malicious code penetration, with code and scripts entering networks from infected or compromised web pages, HTTP represents the bulk of traffic passing through these gateway devices. This approach thus covers all but a tiny fraction of transferred data, as shown by its performance in full-scale appliances, drastically cutting the load on the other protective technologies included in such equipment.

Operating alongside this system, a cloud-based file lookup is also used for files from unclassified sites and for protocols not providing URI information, with much lower demands thanks to the initial URI filtering. As the vast majority of malware carried via non-HTTP protocols consists of simple, static files, these samples can be identified by simple checksums, removing the requirement for complex file data to be gathered, sent to the cloud and checked.

These two overlapping layers of protection can operate via existing universal technology—the DNS lookup system used across the internet—to deliver high-speed responses to cloud lookups. The back-end databases can also monitor lookups coming in, watching for trends and patterns and using the telemetry information gathered to provide additional real-time reactivity to new outbreaks. The style of protection provided perfectly complements the more file-based approach of the full-powered desktop solutions it backs up, and is extendable to a wide range of similarly low-resource, high-throughput devices—even to mobile devices like smart-phones given the plummeting costs of connectivity.

Together, the two layers provide a level of protection at least 95% as accurate as in-depth, resource-hungry per-file inspection, and in many cases improving on detection rates of file-scanning techniques thanks to skirting the problem of server-side polymorphism. Reaction times for new attacks are extremely fast, resource consumption is minimal and physical updating of firmware is drastically reduced. All in all, this hybrid approach ticks all the boxes, fitting the constraints of lightweight environments without compromising on protection.

Boston, USA | Oxford, UK
© Copyright 2009. Sophos Plc

*All registered trademarks and copyrights are understood and recognized by Sophos.
No part of this publication may be reproduced, stored in a retrieval system, or transmitted by any
form or by any means without the prior written permission of the publishers.*

SOPHOS
WWW.SOPHOS.COM