Network-Based Malware Detection 2.0: Assessing Scale, Accuracy and Deployment

Version 1.5

Released: July 10, 2013

Securosis, L.L.C. 515 E. Carefree Highway Suite #766 Phoenix, AZ 85085 **T** 602-412-3051

info@securosis.com www.securosis.com

Author's Note

The content in this report was developed independently of any sponsors. It is based on material originally posted on the Securosis blog, but has been enhanced, reviewed, and professionally edited.

Special thanks to Chris Pepper for editing and content support.

Licensed by Palo Alto Networks

Palo Alto Networks® is the network security company. Its innovative platform allows enterprises, service providers, and government entities to secure their networks and safely enable the increasingly complex and rapidly growing number of applications running on their networks. The core of Palo Alto Networks platform is its Next-Generation Firewall, which delivers application, user, and content visibility and control integrated within the firewall through its proprietary hardware and software architecture. Palo Alto Networks products and services can address a broad range of network security requirements, from the data center to the network perimeter, as well as the distributed enterprise, which includes branch offices and a growing number of mobile devices. Palo Alto Networks products are used by more than 12,500 customers in over 100 countries. For more information, visit www.paloaltonetworks.com.

Copyright

Table of Contents

Advanced Attackers Take No Prisoners

It was simpler back then. You know, back in the olden days of 2003. Viruses were predictable, your AV vendor provided up to date virus signatures to catch malware, and severe outbreaks like Melissa and SQL Slammer were successful due to brittle operating systems and poor patching practices — not because it was impossible to defend yourself against recent threats. Those days are long gone, under an onslaught of innovative attacks leveraging professional software development tactics and taking advantage of the path of least resistance: generally your employees.

We have written extensively about battling advanced attackers, the top issue facing many security organizations today. From our original Network-based Malware Detection paper, through Evolving Endpoint Malware Detection, and most recently the *Early Warning arc*: Building an Early Warning System, Network-based Threat Intelligence, and Email-based Threat Intelligence. Most recently we took our message about advanced attackers to executives with the CISO's Guide to Advanced Attackers.

Yet attacks and defenses change continually, so as much as we try to write timeless research, sometimes our stuff needs a refresh. The market for detecting advanced malware on the network has seen rapid change over the 18 months since we wrote the first paper. Compounding the changes in attack tactics and control effectiveness, the competition for network-based malware protection has dramatically intensified, and every network security vendor either has introduced a network-based malware detection capability or will soon. This creates a confusing situation for security practitioners who mostly need to keep malware out of their networks, and are uninterested in vendor sniping and trash talking.

Accelerating change and increasing confusion usually indicate it is time to wade in again and revisit findings to ensure you understand the current decision criteria — in this case of detecting malware on your network. So this paper updates our original research to make sure you have the latest insight for your buying decisions.

Gaining Presence with New Targets

Cloppert's Kill Chain is alive and well, so the first order of attacker business is to gain a foothold in your environment by weaponizing and delivering exploits to compromise devices. Following the path of least resistance, it is far more efficient to target employees and get them to click a link they shouldn't. That is not new, but the targets are. Attackers go after the most widely deployed software to reach as many potential victims as possible, for the best chance of success. Previously this led them to unpatched operating system vulnerabilities. With recent versions of Windows OS exploitation has gotten much harder, which is good for us.

Though the attackers don't accept defeat readily, so they went after the next most widely distributed software: browsers. Their initial success compromising browsers forced browser providers to respond aggressively and lock down their software more effectively. Of course we still see edge case problems with older browsers requiring out-of-cycle patches, but browsers have now largely escaped being the path of least resistance. But the action/reaction cycle marches on with attackers shifting their attention to other widely used software — recently Adobe Reader and Java. And once Oracle and Adobe progress there will be new targets. There always are. *The only thing we can count on is that attackers will find new ways to compromise devices.*

The Role of the Perimeter

Once attackers establish a presence in your network via the first compromised device, they move laterally and systematically toward their targets until they achieve their mission. Our best defense is to detect and block malicious software — hopefully *before* it wreaks havoc in your environment. Once malware establishes itself on a device, you can no longer rely on any device-resident defensive controls to stop it. This issue is particularly acute for endpoints. We talk to an increasing number of organizations which basically treat every endpoint as a hostile device. If it isn't already compromised it will be soon enough. They use preemptive measures such as extensive network segmentation to make it harder for attackers to access their critical data. But they want to stop malware from reaching endpoints in the first place.

> We talk to an increasing number of organizations which basically treat every endpoint as a hostile device. If it isn't already compromised it will be soon enough.

There is clear precedent for this approach. Years ago anti-spam technology ran on internal email servers. But blocking technology moved out to the perimeter, and eventually into the cloud, to shift the flood (and bandwidth cost) of bad email as far away from real email systems as possible. We expect a similar shift in the locus of advanced malware protection, from exclusively endpoint-centric to include the perimeter. But that begs the question: how can you detect malware on the perimeter? With a network-based malware detection (NBMD) device, of course.

As we described in our original paper, these devices analyze files passing on the wire **before** they enter your network, and identify questionable files by executing them in a sandbox and observing their behavior.

Insecurity by Obscurity

Traditional anti-virus worked by matching the malware against a list of signatures from known bad files; matches were blocked as viruses. This endpoint-centric blacklist approach worked well until it broke down. Today it is largely ineffective so endpoint protection vendors have shifted to a combination of heuristics, cloud-based repositories, IP and file reputation, and a variety of other intelligence-based mechanisms to identify malware.

> But attackers are smart — they have figured out how to defeat blacklists, reputation, and most other current anti-malware defenses. Every way they can, they make it difficult to detect their attacks — defeating our security with their obscurity.

But attackers are smart — they have figured out how to defeat blacklists, reputation, and most other current anti-malware defenses. They use polymorphic files that change randomly to defeat your blacklist. They hijack system files normally exempted from analysis by anti-malware agents. They obscure communications with command and control networks to escape detection by IP reputation defenses. Every way they can, they make it difficult to detect their attacks — defeating our security with their obscurity.

This has created an industry-wide arms race that continues to get fiercer as attacker sophistication increases. For example, malware kits now check to see whether they are executing in a virtual machine — playing dead (sometimes by delaying execution for hours or days) in virtual environments, waiting for their chance to run *outside* the security sandbox. Virtualization is used heavily to make sandboxing practical, so sandbox-aware malware escapes detection by some NBMD devices. These new innovative malware techniques make the security and accuracy of NBMD devices more important than ever. We need better detection to justify further investment and yet another device on the perimeter.

Compounding the issue, we see no end in sight for the exponential growth in traffic volume and quantity of malware. This imposes a significant scaling requirement on perimeter NBMD equipment — especially because we increasingly expect organizations to deploy NBMD inline to reliably block malicious files. *In the face of acute funding and resource shortages, and the costs of investigation and remediation, it has become even more critical to block as much malware at the edge as possible.* But going inline to enable blocking substantially increases the latency, security, and reliability requirements of these devices. It is always a bad day when an incremental security device knocks down a network or blocks legitimate traffic — as some of you have learned the hard way.

Evolving Network-based Malware Detection

Over the last 18 months attackers have rapidly evolved their tactics to defeat emerging controls such as sandboxing and command & control (C&C) network analysis. As attackers get more sophisticated defenses need to keep pace. So we have focused this paper on tracking the evolution of malware detection capabilities and addressing issues with early NBMD offerings — including scaling, accuracy, and deployment. But first we need to revisit how the technology works. For more detail you can always refer back to our original Network-based Malware Detection paper.

Looking for Bad Behavior

Over the past few years malware detection has moved from file signature matching to isolating behavioral characteristics and defining indicators of compromise. We can no longer judge malware by what it looks like — we need to actually analyze what a file *does* to determine whether it's malicious. We discussed this behavioral analysis in Evolving Endpoint Malware Detection, focusing on how new approaches add context to make malware detection far more effective.

> We can no longer judge malware by what it looks like — we need to actually analyze what a file does to determine whether it's malicious.

Our original paper includes full descriptions of typical indications that a device may be compromised. As a reminder a basic list includes memory corruption/injection/buffer overflows; system file/configuration/registry changes; droppers, downloaders, and other unexpected programs installing code; disabling existing anti-malware protections; and identity and privilege manipulation. Of course this isn't comprehensive — it is just a quick set of indicators to search devices for when you hunt compromises. Additional clues include parent/child process inconsistencies, exploits disguised as patches, keyloggers, and screen grabbing. Of course these behaviors aren't *necessarily* bad — that's why you want to investigate as quickly as possible and determine intent before any outbreak has a chance to spread.

The innovation in the first generation of NBMD devices was implementing a virtual farm of vulnerable devices (sandboxes) in easy-to-deploy appliances, providing a protected and monitored execution environment for risk determination. This enabled organizations to explode malware within the sandbox and observe suspicious behaviors. Depending on the deployment model (inline or out-of-band), the device either fired an alert or could actually block an infected file from reaching its target.

Tracking the C&C Malware Factory

Another aspect of network-based malware detection is identifying egress network traffic which shows patterns typical of communication between compromised devices and their controllers. Advanced attacks start by compromising and gaining control of a device. Then the compromised device establishes contact with its command and control infrastructure to fetch a malware file with specific attack code and instructions on what to attack and when. In Network-based Threat Intelligence, we dug deep into the kinds of indicators you can look for to identify malicious activity on your network, including:

- **Destination:** You can track the destinations of all network requests from your environment, and compare them against known bad places. This requires an IP reputation capability — basically a list of known bad IP addresses. Of course IP reputation can be gamed using web-based proxies, fast flux domains and dynamic DNS, so combining the reputation with DNS analysis to identify likely Domain Generation Algorithms (DGA) helps eliminate false positives.

- **Strange times:** If you see an uncharacteristic pattern or volume of traffic — such as the marketing group suddenly performing SQL queries against engineering databases — it's time to investigate.

- **Applications, file types, contents, and protocols:** You can learn a lot by monitoring all egress traffic for large file transfers, non-standard protocols (typically encapsulated in HTTP or HTTPS), strangely encrypted files, and anything else that seems a bit off… Profiling outbound application traffic using the application awareness capabilities of new network security devices can also provide a baseline to identify "non-normal" communications patterns. These anomalies don't necessarily pinpoint compromise but do warrant further investigation.

- **User profiling:** In addition to traffic analysis, we believe it's time to think a little out of the box and profile your users to identify which applications they use and when. This involves taking a granular baseline of user behavior by monitoring applications and activities on the network, and then identifying potentially anomalous activity by those users to provide a place to begin investigating.

Layers FTW

We focus on network-based malware detection in this paper but we cannot afford to forget the need to protect endpoints. It's not just that NBMD gateways will miss stuff. It's that attackers take a broad view of your environment and look for weak spots — wherever they are. Likewise, you need to similarly take a broad and more importantly, integrated view of your defenses breaking down existing silos between network, endpoint and data center security organizations. To be clear, it's naive to believe you can keep computing devices (endpoints or servers) clean. The protection on the endpoints and controls on the network MUST work together to ensure full protection — both when the device is on the corporate network and when it is off.

> The protection on the endpoints and controls on the network MUST work together to ensure full protection — both when the device is on the corporate network and when it is off.

Threat intelligence plays a critical role in dealing with advanced attackers, making both network and endpoint malware detection capabilities smarter and more effective. You want a bidirectional mechanism so malware indicators found by the network device or in the cloud are accessible to endpoint agents for the greatest chance of *detection before infection*, and vice-versa. Malware found on devices should be shipped up for further analysis, profiling, determination, and ultimately distribution of indicators to protect other devices. We will describe this 2nd Derivative Effect (2DE) later in this paper — how the wisdom of crowds has become key to fighting advanced malware.

You may be one of the few, the lucky, and the targeted. No, it's not a new soap opera — it just means you will see interesting malware attacks first. You will catch some and miss others — and by the time you clean up the mess you will probably know a lot about what the malware does, how, and how to detect it. Earn good corporate karma by helping other organizations by sharing what you found, even if you need to remain anonymous. If you aren't a high-profile target this information sharing model works even better for you, allowing you to benefit from the misfortune of the targeted.

The goal is to increase your chance of catching malware before it invades your environment — or at least preventing it from completing its mission and wreaking havoc on your network. This requires a coordinated effort across the network and devices, leveraging a threat intelligence capability to provide your best opportunity to detect an attack *before* it's too late.

Scaling Network-based Malware Detection

Let’s turn our attention to another challenge for this quickly evolving technology: scalability of analysis.

Much of the scaling problem has to do with the increasing sophistication of attackers and their tools. Even unsophisticated attackers can buy sophisticated malware on the Internet with their Bitcoins. There is a well-developed market for packaged malware, and malware writers are capitalizing on it. Market-based economies are a double-edged sword. And that doesn’t even factor in advanced attackers, who routinely discover and weaponize 0-day attacks using them to gain footholds in victim networks, or polymorphic malware that changes randomly at scale. All together this makes the scalability of malware analysis a top requirement for network-based malware detection.

Even unsophisticated attackers can buy sophisticated malware on the Internet. There is a well-developed market for packaged malware, and malware writers are capitalizing on it.

So why is it hard to scale up? There are a few issues:

1. **Operating systems:** Unless your operating system environment is homogeneous, you need to test each malware sample against every vulnerable operating system in your environment. One-to-many testing means each malware sample must be tested against several virtual machines, each running a different operating system.

2. **VM awareness:** Even better, attackers now build logic into the malware to check whether their malware is executing within a virtual machine. If the malware detects it’s running within a VM, it either goes dormant or waits a couple hours in hopes of escaping detection. So to fully test malware the sandbox needs to let it cook for a while. That means you need to spin up multiple VMs and let them run for a while — very resource intensive. Vendors talk about accelerating the clock within the sandbox to address this concern, but it’s not like malware writers can’t put separate timers into their code to avoid trusting the system clock. So VM awareness remains a problem for sandbox-based detection.

3. **Network impact:** Analyzing malware isn’t just a matter of determining that a file is malicious. You also need to understand how it uses the network to connect to its command and control infrastructure and performs internal reconnaissance to move laterally toward its target. That requires watching the network stack and parsing network traffic patterns on every VM in the sandbox.

4. **Analyze everything:** You cannot limit deep analysis to files that look *obviously* bad based on simple file characteristics (static analysis). With the advanced obfuscation techniques in use today you need to analyze *all* unknown files. Given the number of files entering a typical enterprise network daily, these devices need to be able to scale to tens of thousands of files per day — at least.

As you can see, the computing requirements to fully test inbound files are substantial and growing rapidly. One option for dealing with this problem is to reduce the analysis. You could certainly make a risk-based decision to disregard VM-aware malware. You might decide not to analyze documents or spreadsheets for macros. You may not worry about the network characteristics of malware. These are all legitimate ways to scale network-based malware detection without installing a lot more iron. But each compromise weakens your ability to detect malware. Everything comes back to risk management and tradeoffs. But, for what it’s worth, we believe reducing malware analysis is a bad idea.

Historically the answer to most scaling problems has been to add computing power — generally more and/or bigger boxes. Vendors selling boxes love that answer, of course. Enterprise customers, not so much.

Scaling the Malware Analysis Mountain

Historically the answer to most scaling problems has been to add computing power — generally more and/or bigger boxes. Vendors selling boxes love that answer, of course. Enterprise customers, not so much. Scaling malware detection by adding hardware raises two significant issues. First is cost. Not just the cost of the product, although each box requires a threat update subscription and maintenance. But don’t forget the additional operational cost of managing more devices. Setting and maintaining policies can be challenging; ensuring each device is operational, properly configured, and patched adds more overhead. You need to keep each NBMD device up to date (daily) — new malware indicators appear constantly, and need to be loaded onto each device to stay current.

We have seen this movie before. There was a time organizations ran anti-spam devices within their own networks using expensive enterprise-grade gear. When the volume of spam mushroomed enterprises needed to add devices to analyze all the inbound mail and keep it flowing. This was great for vendors but made customers cranky. The similarities to network-based malware detection are clear. We won't keep you in suspense … the anti-spam story ended in the cloud. Organizations realized they could make scaling *someone else's* problem with managed email security services. So they did, *en masse*. This shifted the burdens of keeping up with the flood of spam and keeping devices operational and current onto providers. We expect a similar end to the NBMD game.

We understand that many organizations have already committed to on-premise devices. They need to figure out how to scale existing infrastructure. This requires central management from the vendor and a clear operational process for updating devices daily. At this point on-premise NBMD devices are mature enough to have decent central management capabilities, allowing users to configure policies and deploy updates throughout the enterprise, or have the devices download updates directly from the vendor. We recommend some human oversight of policy updates because automatic updates have a sad history of bricking devices and stopping traffic flow into the network.

But we expect organizations to increasingly choose cloud-based malware analysis, coordinated with on-premise enforcement devices for collection and blocking. This shifts responsibility for scaling and updating the analysis engine onto the provider. But accountability cannot be outsourced so you need to ensure both detection accuracy (which we will discuss later in the paper) and reasonable turnaround times for a determination on each file. Make sure to build this oversight into your processes.

Another benefit of the cloud approach is easy sharing of malware indicators, so any malicious activity found in any protected customer network can be used by all that vendor's other customers. This offers tremendous leverage, especially to smaller organizations, because security providers see far more malware than their customers. Benefiting from others' misfortune makes good business sense and is the basis for threat intelligence.

Cloud Concerns

As great as this cloud stuff sounds, there are legitimate concerns with cloud-based malware analysis. Let's start with latency. The laws of physics insist it will always take time to ship a malware file up to the cloud, have it analyzed, and then receive the verdict. This creates a potential exploit window which must be managed. It's not practical to hold files at the perimeter until a determination can be made, as it would totally break the Internet user experience. So you'll need to be prepared to clean up the mess later if the file turns out to be malicious.

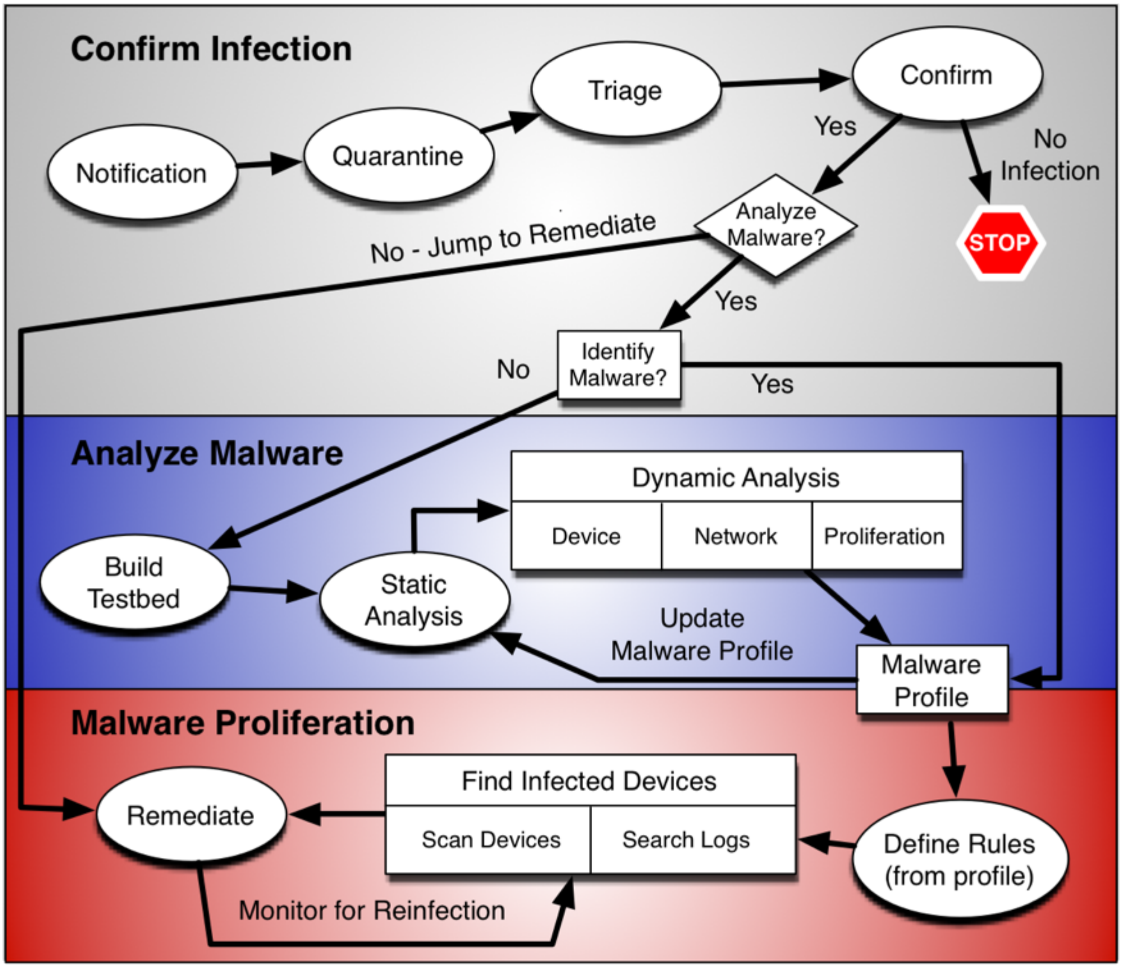
> The laws of physics insist it will always take time to ship a malware file up to the cloud, have it analyzed, and then receive the verdict. This creates a potential exploit window which must be managed.

Another issue is information sharing. Some organizations, including the military and other high-security groups, remain reluctant to share information or send malware beyond organizational boundaries. They may never be comfortable doing the malware analysis in the cloud so they stick to on-premise options. We expect vendors espousing cloud-based approaches to eventually recognize the importance of this use case and offer customer-premise variants of their cloud technology. These malware analysis “private clouds” will be based around a central analysis device, and scale up by adding more hardware to the private cloud. To leverage threat intelligence from other customers they will receive tightly controlled inbound-only updates of new indicators from the latest attacks.

For those vendors pushing an on-premise device, we expect they’ll shift toward a hybrid approach as well, taking advantage of the cloud as appropriate, because many smaller enterprises will find cloud-based services irresistible given their cost and scalability advantages.

The Network’s Place in the Malware Lifecycle

Now let’s dig into the malware detection and analysis lifecycle, for context on where network-based malware analysis fits in and what other controls NBMD devices need to integrate with to protect against advanced threats. We have researched malware analysis exhaustively. The process diagram below comes from our Malware Analysis Quant research.

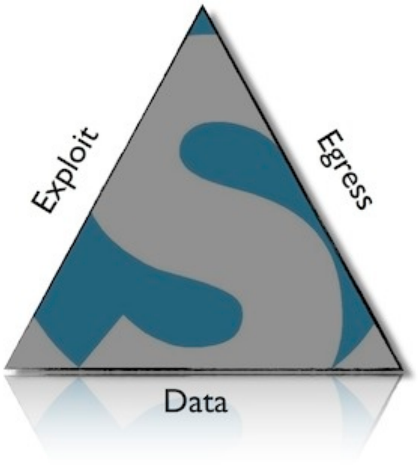


NBMD fits squarely into the analyze malware activity phase — including building the testbed, static analysis, various dynamic analysis tests, and finally packaging everything up into a malware profile.

All these functions occur either on the device or in a cloud-based analysis sandbox. That is why scalability is so important — you need to analyze *every* file that comes in, not just bad ones.

Some other aspects of this lifecycle bear mentioning:

- **Ingress analysis is not enough:** Detecting and blocking malware on the perimeter is a central pillar of the strategy, but no NBMD capability can be 100% accurate and catch everything. You need controls on endpoints, extensive monitoring on internal networks (to detect lateral movement of the attackers), and aggressive egress filtering to stop exfiltration.
- **Intelligence drives accuracy:** Malware and attack tactics evolve quickly, so on-device analysis techniques must as well. This requires significant and sustained investment in threat research and intelligence sharing.



Before we can dig into these two points, we need to point out some other relevant research for additional context. The Securosis Data Breach Triangle shows the opportunities to interrupt a data breach. You can protect the data (very difficult), detect and stop the exploit, or catch the data with egress filtering. Success at any of these points stops a breach, and putting all your eggs in one basket is unwise, so work on all three.

For specifics on detecting and stopping exploits refer to our CISO's Guide to Advanced Attackers — particularly Breaking the Kill Chain on stopping advanced attackers. Remember — even if a device is compromised, unless critical data is exfiltrated, it's not a breach. The best case is to detect the malware *before* it causes any harm — NBMD is very useful for this — but you also need to rely heavily on your incident response process to ensure you contain the damage of the inevitable incidents.

Ingress Accuracy

Accuracy of detection remains a critical success criteria for NBMD. A false positive — incorrectly flagging a file as malware — disrupts work and wastes resources investigating a malware outbreak that never happened. False negatives — missing malware and letting it through — are at least as bad, since you now depend on your other controls to prevent an infection.

So how can you verify the accuracy of an NBMD device, especially since there is no accepted detection accuracy benchmark? You'll need to do some homework on your own. Start by asking potential vendors tough questions to understand their threat intelligence and threat research capabilities. Read their threat research reports to figure out whether they are on the leading edge of research, or just fast followers piggybacking on other companies' research innovations.

> A false positive — incorrectly flagging a file as malware — disrupts work and wastes resources investigating a malware outbreak that never happened. You need to avoid these, so put a premium on accuracy. False negatives — missing malware and letting it through — are at least as bad.

Malware research is critical because it provides the content that drives all malware analysis, both on devices and in the cloud. Understand the depth and breadth of vendor research capabilities. Dig deep to understand how many researchers they have focused on malware analysis. Learn how they aggregate the millions of malware samples in the wild to isolate patterns using fancy terms like big data analytics. Study and understand how they turn that research into detection rules and on-device tests.

Also understand how the vendor shares information with the broader security research community. No single company can do it all, so learn how they collaborate with other groups and what alternative data sources they leverage to broaden their access to attack data.

Make sure you check out lab tests of devices to compare accuracy. These tests are all flawed — it is *extremely* difficult to accurately model a real-world environment using live ammunition (malware) — but they can provide useful apples-to-apples comparison of the devices.

As part of the proof of concept test we recommend you route ingress traffic through 2 or 3 devices in monitoring mode, comparing their accuracy and scalability on real traffic. That should give you a good indication of how well the devices would perform in your environment.

The Second Derivative

Finally, leverage the 2nd Derivative Effect (2DE) of malware analysis through your NBMD vendor. When new malware is found, profiled, and determined to be bad, your vendor has a golden opportunity to inoculate all its customers. This involves the malware analysis system (either on-premise or within the vendor's cloud) storing indicators, behaviors, and rules to identify and block the malware in the vendor's cloud repository; then distributing that intelligence back out to all customer NBMD devices. Keep in mind that polymorphic malware can change specific indicators as well, so to keep pace with the malware writers it's critical for the NBMD device to derive consistent indicators on all of the variants of a malware family.

Sharing these indicators/behaviors amongst all the devices is the network effect in action. The more devices in the network, the more likely the malware will show up somewhere to be profiled before it targets you or any other customer. You can't always rely on the 2DE, but it's as good a plan as any.

It sucks to be the first company infected — you miss the attack on its way in. But at least *everyone else* in the network benefits from your misfortune. The 2DE feedback cycle requires extensive automation, with checks and balances to reduce bad updates and to accelerate distribution of new indicators to devices in the field. Ask many questions and understand how the vendor harnesses the network effect to protect your environment.

It sucks to be the first company infected — you miss the attack on its way in. But at least everyone else in the network benefits from your misfortune.

Deployment Considerations

Taking full advantage of NBMD for detection of advanced malware requires a number of critical decisions. You need to determine how the cloud fits into your plans. Early NBMD devices evaluated malware within the device (on-box sandbox), but recent advances and new offerings have moved some or all malware analysis to cloud compute farms. You also need to figure out whether to deploy the device inline to block malware before it gets in. Blocking whatever you can sounds obvious but there are trade-offs — as always. Let's dig into these decisions, and the pros and cons of each.

To Cloud or Not to Cloud?

On-device vs. cloud sandboxing has become one of those religious battlegrounds vendors use to differentiate their offerings from each other. Each company offering NBMD has a 70-slide deck to blow holes in their competition's approach. We have no use for technology religion, so let's take an objective look at the options. The biggest advantage to on-device sandboxing is reduced latency — you don't need to ship the file anywhere so you may get a quicker verdict. Although this assumes the on-premise NBMD device can keep pace with the speed of a cloud-based analysis farm (which has access to elastic computing power). Remember that latency involves not just the time to send the file for analysis, but also the time to analyze the file as well.

> You need to evaluate every file that comes in through every ingress point, unless you can immediately tell that it's bad from a file hash. That require an analysis capability on every Internet connection to avoid missing something.

But there are real issues with on-device analysis, starting with scalability. You need to evaluate every file that comes in through every ingress point, unless you can immediately tell that it's bad from a file hash. That requires an analysis capability on every Internet connection to avoid missing something. Depending on your network architecture this may be a serious problem. Centralized ingress and egress in a small number of locations makes it easy, while the on-device approach can be cost prohibitive for distributed networks with many ingress points.

We discussed the 2nd Derivative Effect (2DE) earlier, whereby customers benefit from the network effect of working with a vendor which analyzes a large quantity of malware across many customers. The 2DE is also affected by the choice of where to analyze malware. With on-device analysis malware determinations must be sent up to the central distribution point, normalized, and de-duped; then the indicators and tests need to be distributed to all on-premise devices — a multi-step process. Those additional steps extend the window of exposure. On the other hand cloud analysis effectively provides a central repository for all file hashes, indicators, and testing — which significantly simplifies data management and reduces the need to keep tens of thousands of devices across all the vendor's customers.

As we mentioned earlier, we expect cloud-based malware analysis to prevail over time. Your internal analysis may well determine that reduced latency is more important than cost, scalability, and management overhead — and we're fine with that. Just make sure you understand the trade-offs before making a decision.

Inline vs. out-of-band

The next deployment crossroads is where NMBD devices sit in the network flow. Is the device deployed inline so it can block traffic? Or will it be used more as a monitor, inspecting traffic via a span port and sending alerts when malware flies by (potentially infecting devices)? We see the vast majority of NBMD devices currently deployed out-of-band, with organizations working through alerts as quickly as they can. They have determined that delaying the delivery of files during analysis (whether on-box or in the cloud) risks alienating employees, so they haven't fought to deploy inline at this point.

Another issue with out-of-band deployment is working through the alerts. Each alert requires someone to do something, meaning the alert must be investigated, and malware must be identified and remediated quickly enough to contain any damage. Depending on the staff (or lack thereof) you devote to working through these alerts, you might be cleaning up a mess even when the NBMD device successfully flags malware. That has serious ramifications for the NMBD value proposition, and is another reason to use automatic real-time blocking ASAP.

If the NBMD device you championed goes down and fails closed — blocking everything — you may as well start working on your resume.

Finally, we need to mention the importance of being able to monitor all protocols for malware activity. If the NBMD (perhaps as part of a web filter or standalone device) is only looking at traditional web protocols (Ports 80/443), this provides an exposure for attackers to hide malware in other protocols like SMTP, telnet or possibly FTP.

When you step back to think about it, why *wouldn't* you go inline to block the malware on the perimeter before it can infect anything? Isn't that the whole point of NBMD? Kinda-sorta, but inline deployment is a high wire act. Block the wrong file or break a web app and there is hell to pay. If the NBMD device you championed goes down and fails closed — blocking everything — you may as well start working on your resume. That's why most folks deploy NBMD out-of-band for quite a while, until they feel confident it won't break anything important.

We expect NBMD to eventually run within the *perimeter security gateway*. That is our term for a single box that encompasses NGFW, NGIPS, web filter, and other capabilities. Obviously the perimeter security gateway is inline, so NMBD *will* end up deployed inline. But NBMD need not be deployed in blocking mode, it can alert (and not block) even when inline, offering much more flexibility on what to block and what to alert on. Deploying the device inline offers the best of both worlds.

The Egress Factor

This paper focuses on the detection part of malware response. But we need to at least touch on preventative techniques to ensure critical data doesn't leave your network, even if malware *does* penetrate your perimeter. Per the Securosis Data Breach Triangle, breaking the egress leg and stopping exfiltration means there is no breach. That is simple to say but hard to do. Compounding the issue, most outbound network traffic is now encapsulated in port 80 or 443, and we continually see new exfiltration mechanisms. We have seen tampering with consumer storage protocols (Google Drive & Dropbox) to slip files out of a network, as well as exfiltration 140 characters at a time through Twitter. Attackers can be pretty slick to make your job harder.

> Ultimately you should be able to take an egress default deny posture. This allows authorized applications to send data out of your network while everything else is blocked.

So what to do? We continue to recommend aggressive egress filtering on your perimeter and blocking unknown applications and file types. If you cannot identify an application in the outbound stream, block it. If you cannot see into a file because it is encrypted or an unknown type, block it. This requires NGFW-type application inspection and classification capabilities, as well as simple web filtering functionality. Ultimately you should be able to take an egress *default deny* posture. This allows authorized applications to send data out of your network while everything else is blocked.

As we discussed in Network-based Threat Intelligence, you can block traffic to known bad websites as well. Using a list of known malware sites and other places you don't want employees browsing, set your egress filter to block traffic to your IP blacklist. This approach still has all the normal blacklist limitations so be selective about what you block vs. merely alert on, and understand that keeping the list current will be challenging. Combining an whitelist of applications approved for outbound traffic with a blacklist of bad locations can increase the effectiveness of egress filtering to break the attack chain and stop exfiltration while retaining flexibility.

Summary

It is common sense to handle attacks as far away from your devices and critical data as possible. So the security industry has made concerted efforts to move protection away from endpoints and data centers, into the network perimeter and cloud-based services. We saw this in email security years ago and see it in malware detection today.

Network-based Malware Detection (NBMD) is one of the hottest product areas in security right now, as improving detection techniques and evolving architectures have addressed scalability and accuracy problems with the first generation of these devices, and competitive offerings from network security players have improved product quality and effectiveness.

The biggest area of innovation has been the rapid maturity of cloud-based malware analysis sandboxes, which are now tightly integrated into NBMD enforcement devices in the perimeter. This enables organizations to scale up their environments cost-effectively and leverage the Second Derivative Effect (2DE) to benefit from the breadth of malware seen by vendors.

We have also seen most NBMD deployments start, and many remain out-of-band due to the risk of blocking the wrong traffic. But over time we expect NBMD capabilities to be subsumed into the inline perimeter security gateway, giving customers the flexibility to block stuff that's obviously bad at the gateway and alert on questionable files that warrant human investigation.

Inevitably NBMD devices will miss and malware will make it through your perimeter. This eventuality requires integration between the NBMD device and the other security controls in your environment (including endpoint protection) to search for and block indicators discovered through incident response. Finally, comprehensive malware defense requires egress filtering to break the third leg of the Data Breach Triangle: exfiltration. Even when devices are compromised, this gives you another chance to avoid a breach.

If you have any questions on this topic, or want to discuss your situation specifically, feel free to send us a note at info@securosis.com or ask via the Securosis Nexus (http://nexus.securosis.com/).

About the Analyst

Mike Rothman, Analyst/President

Mike's bold perspectives and irreverent style are invaluable as companies determine effective strategies to grapple with the dynamic security threatscape. Mike specializes in the sexy aspects of security — such as protecting networks and endpoints, security management, and compliance. Mike is one of the most sought-after speakers and commentators in the security business, and brings a deep background in information security. After 20 years in and around security, he's one of the guys who "knows where the bodies are buried" in the space.

Starting his career as a programmer and networking consultant, Mike joined META Group in 1993 and spearheaded META's initial foray into information security research. Mike left META in 1998 to found SHYM Technology, a pioneer in the PKI software market, and then held executive roles at CipherTrust and TruSecure. After getting fed up with vendor life, Mike started Security Incite in 2006 to provide a voice of reason in an over-hyped yet underwhelming security industry. After taking a short detour as Senior VP, Strategy at eIQnetworks to chase shiny objects in security and compliance management, Mike joined Securosis with a rejuvenated cynicism about the state of security and what it takes to survive as a security professional.

Mike published The Pragmatic CSO <http://www.pragmaticcso.com/> in 2007 to introduce technically oriented security professionals to the nuances of what is required to be a senior security professional. He also possesses a very expensive engineering degree in Operations Research and Industrial Engineering from Cornell University. His folks are overjoyed that he uses literally zero percent of his education on a daily basis. He can be reached at mrothman (at) securosis (dot) com.

About Securosis

Securosis, LLC is an independent research and analysis firm dedicated to thought leadership, objectivity, and transparency. Our analysts have all held executive level positions and are dedicated to providing high-value, pragmatic advisory services. Our services include:

- **The Securosis Nexus**: The Securosis Nexus is an online environment to help you get your job done better and faster. It provides pragmatic research on security topics that tells you exactly what you need to know, backed with industry-leading expert advice to answer your questions. The Nexus was designed to be fast and easy to use, and to get you the information you need as quickly as possible. Access it at <https://nexus.securosis.com/>.

- **Primary research publishing**: We currently release the vast majority of our research for free through our blog, and archive it in our Research Library. Most of these research documents can be sponsored for distribution on an annual basis. All published materials and presentations meet our strict objectivity requirements and conform to our Totally Transparent Research policy.

- **Research products and strategic advisory services for end users**: Securosis will be introducing a line of research products and inquiry-based subscription services designed to assist end user organizations in accelerating project and program success. Additional advisory projects are also available, including product selection assistance, technology and architecture strategy, education, security management evaluations, and risk assessment.

- **Retainer services for vendors**: Although we will accept briefings from anyone, some vendors opt for a tighter, ongoing relationship. We offer a number of flexible retainer packages. Services available as part of a retainer package include market and product analysis and strategy, technology guidance, product evaluation, and merger and acquisition assessment. Even with paid clients, we maintain our strict objectivity and confidentiality requirements. More information on our retainer services (PDF) is available.

- **External speaking and editorial**: Securosis analysts frequently speak at industry events, give online presentations, and write and/or speak for a variety of publications and media.

- **Other expert services**: Securosis analysts are available for other services as well, including Strategic Advisory Days, Strategy Consulting engagements, and Investor Services. These tend to be customized to meet a client's particular requirements.

Our clients range from stealth startups to some of the best known technology vendors and end users. Clients include large financial institutions, institutional investors, mid-sized enterprises, and major security vendors.

Additionally, Securosis partners with security testing labs to provide unique product evaluations that combine in-depth technical analysis with high-level product, architecture, and market analysis. For more information about Securosis, visit our website: <http://securosis.com/>.