Whitepaper

Nine Metadata Use Cases: How to Use Metadata to Make Data-Driven Decisions

We have reached that point. The one where there is so much data on the network—in terms of volume, variety of data types, and speed at which it moves—that detecting good traffic from bad is not only costly, but, with the high signal-to-noise ratio, almost impossible for most security tools to handle. Attackers know this, and they understand how easy it is to go unheard—and unseen—in all that noise.

Unfortunately, we have also reached the point where there is too little time and too little compute resources to efficiently correlate all the information required to build relevant context to make accurate predictions on potential security threats. This includes more than the attacks coming from outside the network; security teams can also easily miss insider abuse hidden within the noise.

No doubt, today's advanced security information and event management systems (SIEMs) can help. They are valuable correlation engines capable of ingesting a great many different things. However, for all the promise of Big Data, it remains difficult to manage enough compute across all the required and varied data sets to draw inferences about whether logged or observed system events are good or bad.

To create context about an event, certain information is needed. For instance, it is important to know the IP of a machine in question; which user is currently logged on it and which user had been on it historically; what website was visited; what content delivery network (CDN) was used; what sort of certificate was sent when SSL began; who signed it; and more. And this is all before any inferences can be made.

A SIEM must then spend valuable cycles stitching and grinding together this basic information to produce a tiny not-so-valuable drop of information. With time, it will, of course, produce more and more drops until, eventually, it can begin to correlate across them all to produce something of value.

There must be—and there is—a better way to take all of those different types of structured data and:

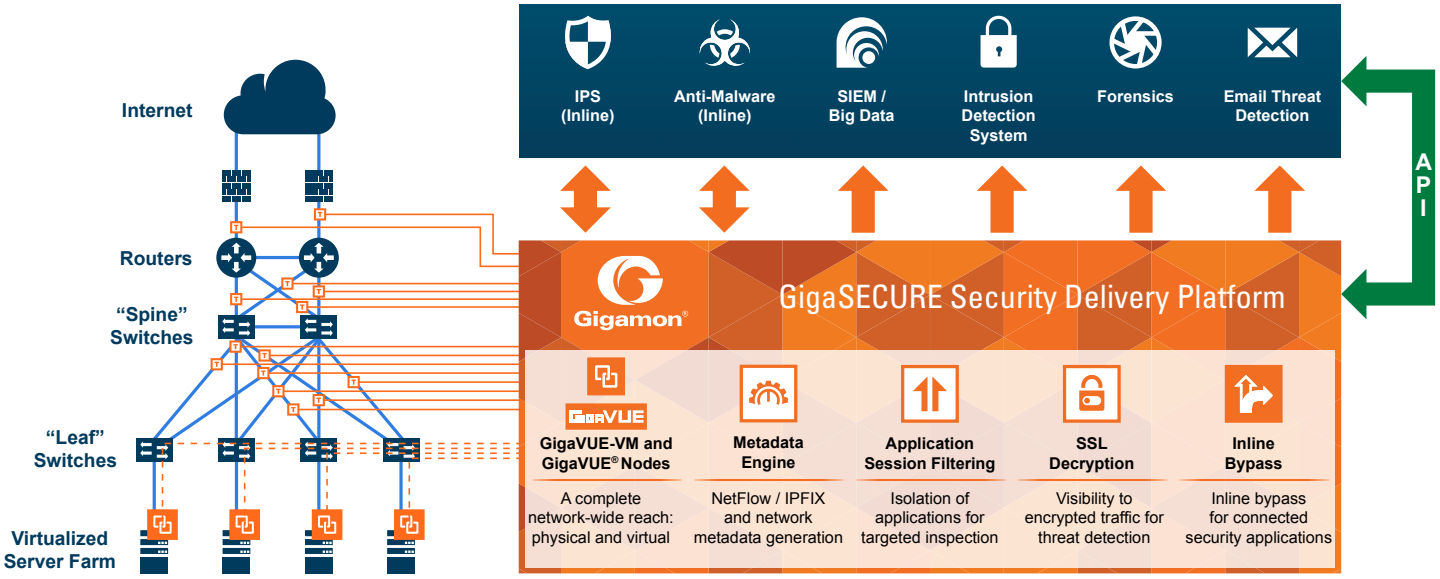
- go to them with a specific problem,
- pull out the data that is relevant to that problem,
- create a single summary record without using any compute cycles,
- send thousands of those now highly enriched summary records to a SIEM,
- and then—and only then—burn compute cycles to analyze those records.

It's called metadata. And today, Gigamon can give a SIEM and next-generation user behavior analytics (UBA) tools the valuable drops—the metadata—before the analysis grinding, let alone correlation, even begins.

The Gigamon Solution: GigaSECURE Security Delivery Platform

The GigaSECURE® Security Delivery Platform connects into the network, both physical and virtual. With the ability to see 100 percent of every traffic packet sent and received, it can be configured to deliver relevant network traffic—in the form of unsampled IPFIX records, URL/URI information, SIP request information, HTTP response codes, DNS queries, or SSL certificate information—to SIEMs (for real-time situational awareness and anomaly detection) or collectors (for rapid forensic analysis after an incident).

Often, these elements come from different sources and in different formats and, if they are able to be collected at all, they need to be rejoined when they arrive at a security ingestion point such as Splunk or a Hadoop stack. Rather than waste compute cycles trying to correlate everything at that point, GigaSECURE can summarize the interesting parts of network conversations into a single metadata record that can be processed with like records to form context, in real time and with less disk or CPU.



GigaSECURE Security Delivery Platform: Key components

Gigamon's appliances are able to send un-sampled NetFlow to collectors, storage, and SIEMS. With our latest release, if you are consuming NetFlow IPFIX, we are sending more than just standard information about the traffic.

In RFC 7012 for IPFIX, enterprises are able to register private SMI Network Management numbers at IANA. Gigamon's IANA SMI number is 26866. With it, GigaSECURE can send custom templates with extensions that include URL information, HTTP return code, DNS query and response information, SSL certificate information, and other metadata elements in addition to the rich flow record information such as source and destination IP addresses, total bytes of information exchanged, application port number, etc.

Useful Metadata Use Cases include:

1. URL – Uncover Malicious Sites or Undesired Behavior—Even When Traffic Does Not Go through a Proxy
2. URL – Get Visibility into Internal Applications Traffic That Does Not Traverse a Proxy
3. URL – Avoid Proxy Performance and Latency Issues
4. URL – Enforce Policy and Appropriate Controls—Even with Presence of Unknown HTTP/S Traffic
5. DNS – Solve the DNS Recursive Lookup Dilemma and Match Endpoints to DNS Requests
6. DNS – Detect DNS Hijacking and Credential Harvesting Attacks
7. DNS – Detect Tunnels and Prevent Bypass of Security Controls
8. HTTP – Use HTTP Return Codes to Detect Malware on the Network
9. SSL – Uncover Malware Hidden in Encrypted Traffic Using SSL Certificate Information

Use Case One (URL):

Uncover Malicious Sites or Undesired Behavior—Even When Traffic Does Not Go through a Proxy

Understanding what is connecting where on the network is a fundamental building block for creating the context needed to conclude whether a connection is good or bad. For this reason, analyzing proxy logs and drawing correlations between normal and anomalous behavior on the network is standard practice.

Traditionally, an organization might gather information on who its users are, what computers they are using, and what departments they belong to. Next, it would look to establish what sites might be malicious or uncover undesired behavior such as an employee viewing inappropriate adult content, working with competitors, or simply spending too much time on non-job-related Web browsing.

This approach, however, would not work as well when traffic going to the Internet is not behind a proxy. For example, when the software using the Internet was not written to be proxy aware; when a machine is at a remote site and without a proxy; or when the organization has decided to forgo a proxy because it is either underpowered or would add latency.

Without proxy logs, organizations have limited access to URLs and might miss URI—which could be the difference between knowing what's happening on the network and being able to catch and fix an issue in real time versus discovering an issue after a serious event has occurred and it is too late to take action.

For instance, think about if a new manager were to join an organization and, suddenly, his group starts to become much less productive. The department head might come to the executive committee and request information about what is going wrong. Even if conversations are had, discontent may remain. And what if an employee exodus were to ensue?

What other ways could an organization begin to assess what is happening? How about delving into whether or not there have been changes in workflow? Or finding out if the manager implemented a new tool or removed an old one? If an organization can see these things on the network, it can start to uncover patterns and deduce what's going on faster and more effectively.

One way to solve the problem is to turn huge volumes of data into useful, actionable information. Assuming an organization has proxy logs it can combine and analyze with other unstructured data (e.g., AD DHCP requests; calendar invites to lunch in .PSTs; badge reader data; DNS; etc.), it stands a chance of figuring out what is happening.

With GigaSECURE, even if an organization is without proxy traffic, it can still begin to pinpoint issues.

Once the unstructured data is in the system, an organization can query for specifics–from IP addresses to usernames to URLs to the time—and then spit out a single specific record per user and get a more manageable, visual "graph" of what is going on over a given length of time.

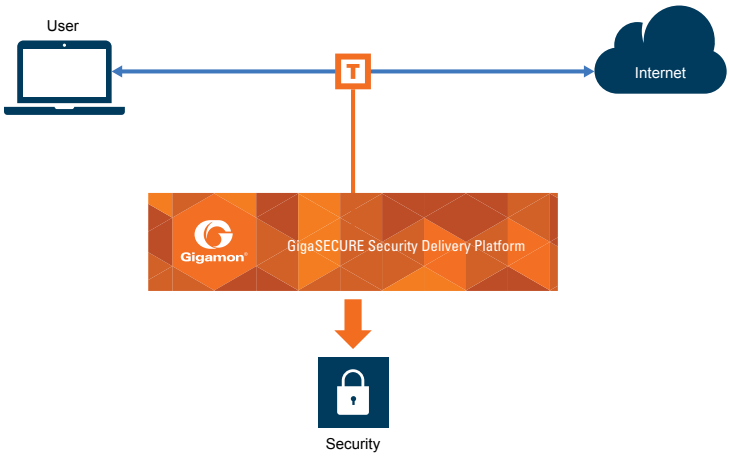


Figure 1: Security without a proxy

At the end of the day, a CISO's job isn't just about stopping advanced persistent threats (APTs) and defending against nation states, it is about keeping good talent on board and preventing employees from performing actions like plugging in random infected thumb drives into domain controls that put the company at risk or sending confidential files because no one told them otherwise.

If organizations can have more data points alerting them to when something may be amiss, they have a better chance at making a positive change before a costly cyber incident or, even, employee departure.

Use Case Two (URL):

Get Visibility into Internal Applications Traffic That Does Not Traverse a Proxy

Traffic that uses internal applications (e.g., airport ticketing kiosks, branch office banking software) may not always traverse a proxy. And as IoT devices become more pervasive, it will also become more challenging to understand where they are connecting in a way that is easy to manage—especially when not every household or factory floor is likely to run proxies to monitor them as they traverse north out of the network, much less east-west within.

An obvious use case is the ability to look for insider threat. Why is a system that shouldn't be accessing internal tools suddenly connecting to those systems? Which user is logged in and why is he accessing the unauthorized systems?

Some organizations have more than 2,000 unique applications and tools. Daily, there could be terabytes of log data about their usage; and sending that information in log format could burden the network. With GigaSECURE attached to the network, organizations can collect metadata and send it out of band to tools that perform analysis.

Imagine an airport kiosk, where airline employees are checking in customers, printing out boarding passes. Or a bank branch office, where tellers are viewing account information, transferring money, resetting passwords, etc. To perform these operations, the employees are connecting to customized, purpose-built tools owned by, respectively, the airline or financial institution. Within any single corporation, there can be tens of thousands of these types of proprietary tools; many of which handle sensitive, confidential data, but are also legacy and not designed for adequate logging.

By contrast, on the network, an organization can see URLs in the IPFIX records and, with a single change to the GigaSECURE platform, it can send that information to a SIEM. If strange behavior is noticed, it can make a correlation faster and more efficiently than trying to correlate logs from the endpoint or application server. With pervasive visibility of network traffic, an organization can send those URLs to detection machines to determine—based on company rules—if what a person is doing is good or bad.

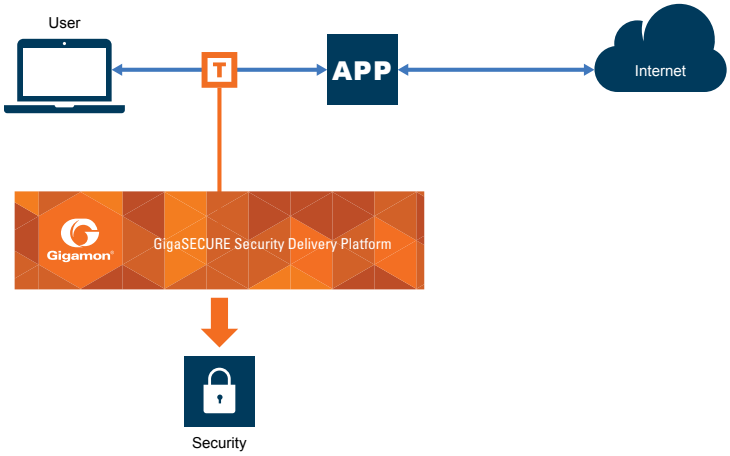


Figure 2: Visibility into internal application traffic without logs

GigaSECURE gives companies visibility into actions on the network so they can see data in use (e.g., a teller connecting to a backend system where money can be moved) without relying on attempting to retrieve logs off a backend system or endpoint in order to make a correlation. With the IP and URL addresses, for instance, they can more easily make correlations and predict bad behavior on their networks; something that they might not have been able to see before, and that they now can even though the traffic is not proxied or they are not able to retrieve the logs.

Use Case Three (URL):

Avoid Proxy Performance and Latency Issues

A proxy engine is designed for data plane work in order to service Web connections. When an organization tries to assign it the additional responsibility of logging connections, performance can suffer and users will complain that websites are not loading quickly enough. This type of latency can become particularly troublesome within organizations, such as news agencies, where milliseconds are crucial.

When it is necessary to understand what URLs are being accessed, the proxy must move up from the data plane onto the control plane to send logs to SIEMS. This can create a variety of issues. In addition to causing pages to load more slowly, it can fill up disk space and cause CPU to spike.

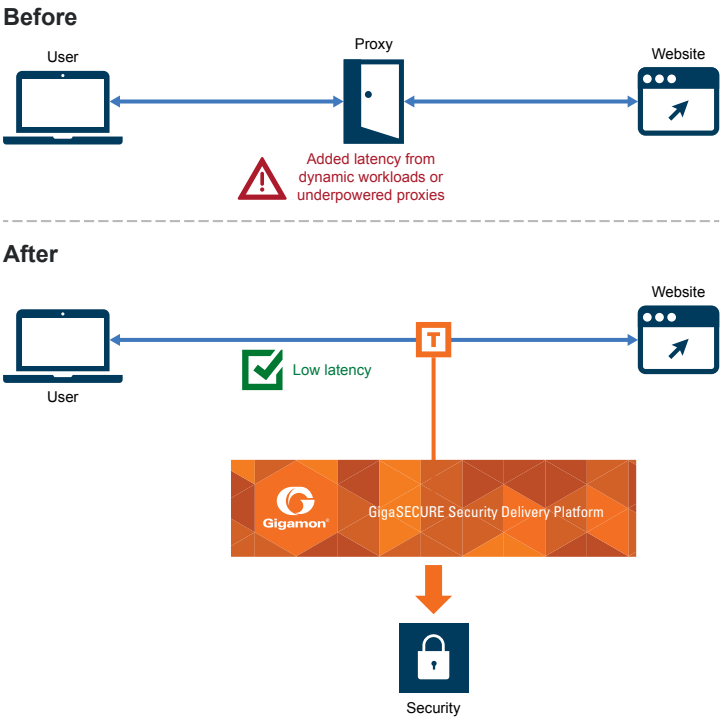


Figure 3: Eliminating proxy latency or power issues

By using GigaSECURE to send IPFIX templates that contain URL and source IP information, not only can organizations trace which sites employees are visiting without any overhead, but they can also correlate threat feeds and indicators of compromise (IoC) with URLs on the network to determine if a machine is breached. Or, they can use the same data to perform statistical analysis about usage to spot unusual behavior and detect breaches. For example, it can reveal that a system administrator in charge of file servers is connecting to URLs that manage payroll—something an organization might want to investigate further.

Use Case Four (URL):

Enforce Policy and Appropriate Controls—Even with the Presence of Unknown HTTP/STraffic

The ability to manage and share information both inside and outside of a company is important. But what would happen if a Web server were installed on a company's network that the security team was not aware of? It happens all the time—because administrative credentials are not needed to run a Web server on an ephemeral port.

Simply detecting HTTP/S traffic—regardless of the port it's running—is a feature that ships and can be enabled in GigaSECURE. The ability to spot HTTP/S traffic running on non-standard ports can tip off a broken workflow, process, or worse.

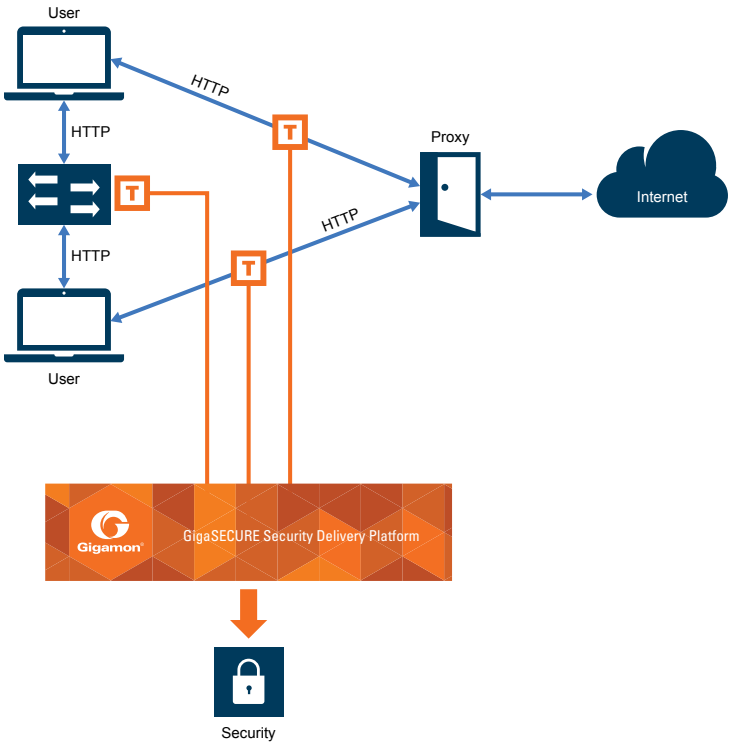


Figure 4: Detecting HTTP/S traffic regardless of port

In some cases, organizations don't expect to see HTTP traffic. Its presence alone could be an indicator of compromise. Or, even if it is simply a new system the organization didn't know about, this type of metadata awareness provides the opportunity to enforce policy and ensure that appropriate controls are put in place.

Use Case Five (DNS):

Solve the DNS Recursive Lookup Dilemma and Match Endpoints to DNS Requests

Domain controllers are Windows machines that run a directory, authentication services, Domain Name System (DNS), and LDAP. When users log into a PC or Mac in a corporate environment, they usually authenticate their username and password against a domain controller, which keeps a copy of their hash, gets them logged in, and decides what files they can see. In short, it manages all permissions.

Now for the more problematic part.

When someone asks for a certain website (e.g., www.gigamon.com), the domain controller looks up the IP address for gigamon.com. But how? In many companies, looking up www.gigamon.com doesn't actually mean the domain controller goes out to the Internet and asks the root server who the authoritative name server is for www.gigamon.com. It almost never works that way. Domain controllers—because they control authentication, passwords, access/right management—are some of the most valuable machines on a network. Therefore, they are also some of your most secure boxes and may never touch the Internet directly.

For this reason, DNS is designed to use something called recursion when making a lookup. This means that when someone asks for www.gigamon.com, the request goes to the DNS server configured on his PC or Mac, which is almost always the domain controller.

In most companies, the next hop for recursion is an Internet-facing server—often purpose-built to do only DNS (e.g., an open source DNS server like BIND or DJBDNS). The domain controller will connect to a DMZ, hit the Internet-connected machine, which, in turn, goes to a root server and asks who the authoritative name server for Gigamon is. It then passes that back to the DNS server to look up the record for www.gigamon.com.

If an organization has thousands of PCs and the domain controller is the first place where DNS is being asked and the domain controller does not directly connect to the Internet-facing server, when looking in the DNS server logs connecting to the Internet, the requesting IP will be the IP of the domain controller. Security analysts need the IP of the computer making the request, but all they can see is what domain was requested and not who made the request.

Until very recently, Microsoft did not design its domain controllers to easily provide DNS logs. Because of this, the only way to get logs is from the Internet-facing server.

For all the thousands of users behind a domain controller and all the tens of thousands or millions of domains requested in a day, the log file would show the DNS server as the only one requesting all of those different URLs. When an organization wants to look into a URL like www.gigamon.com or, perhaps, www.illegalactivities.com, it cannot tell who is going where because the only "person" making requests is the domain controller.

That's the DNS dilemma. Zero visibility.

Metadata can bring light to the DNS darkness, and makes it possible to match endpoints to DNS requests. With GigaSECURE on the network between domain controllers and users, it is possible to pull DNS traffic right off the wire—without putting an agent on a box, without doing any configuration changes.

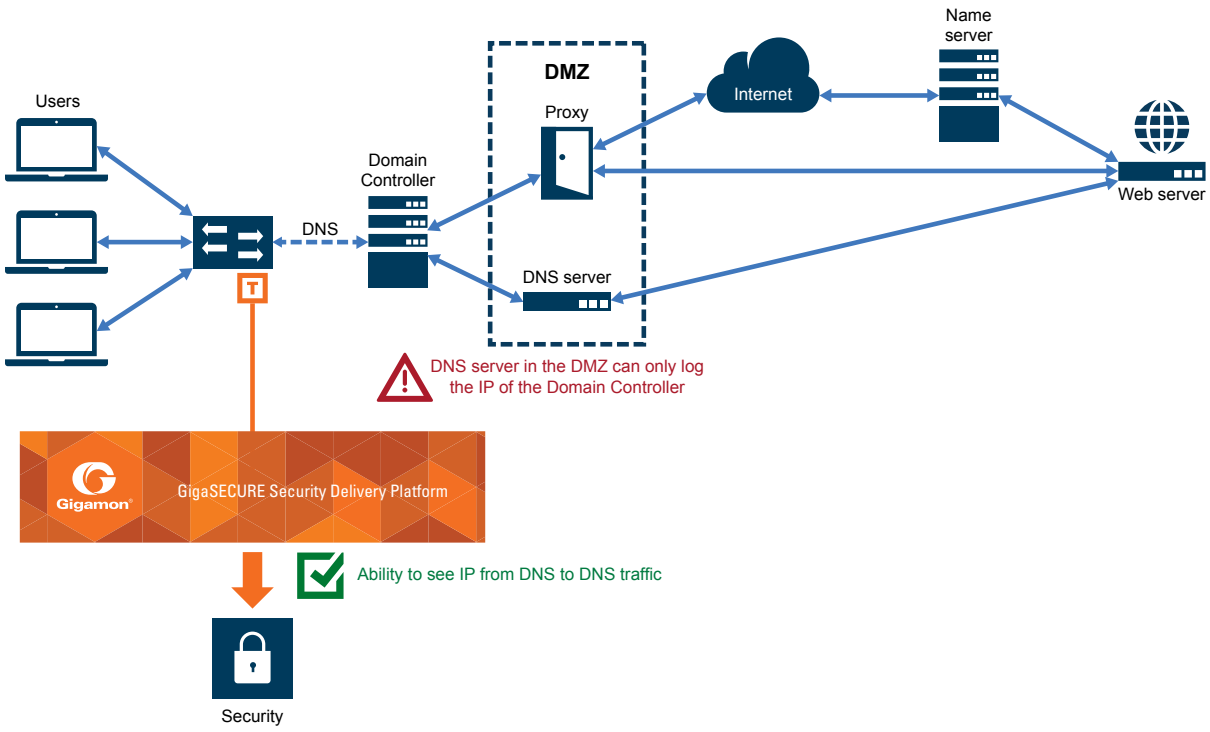


Figure 5: Matching endpoints to DNS requests

More specifically, in environments where DNS goes through several layers of recursion before connecting out of the network, GigaSECURE can be configured to send information about the DNS request to an IPFIX collector or SIEM. This is particularly useful when the endpoint is using Active Directory for DNS requests. The traffic is presented with an easy tag and the endpoint can be matched to DNS request.

Use Case Six (DNS):

Detect DNS Hijacking and Credential Harvesting Attacks

At times, a DNS server cannot send any logs back to a SIEM, for example when a third-party resolver (e.g., Google Public DNS) is configured to perform DNS lookups in the enterprise. In these cases, a common technique for deceiving users about what websites they are visiting is to hijack their local resolver.

With a piece of malware, an attacker can change a local resolver—from the domain controller or the third-party resolver to a hostile one—and redirect users from their desired website to his own. This happens frequently in credential-harvesting attacks. Users think they are logging into, for example, their bank, but are actually entering their usernames and passwords into an attacker's site. The attacker steals their information before passing them through to the bank and, often, victims remain unaware of the data theft.

With GigaSECURE configured to send information to a SIEM, an organization can not only get logs that ensure the correct resolver is being used, but also instantly detect a hijack attack. Using URL and IP metadata, it is easy to correlate and identify when a DNS resolver's IP address has been altered.

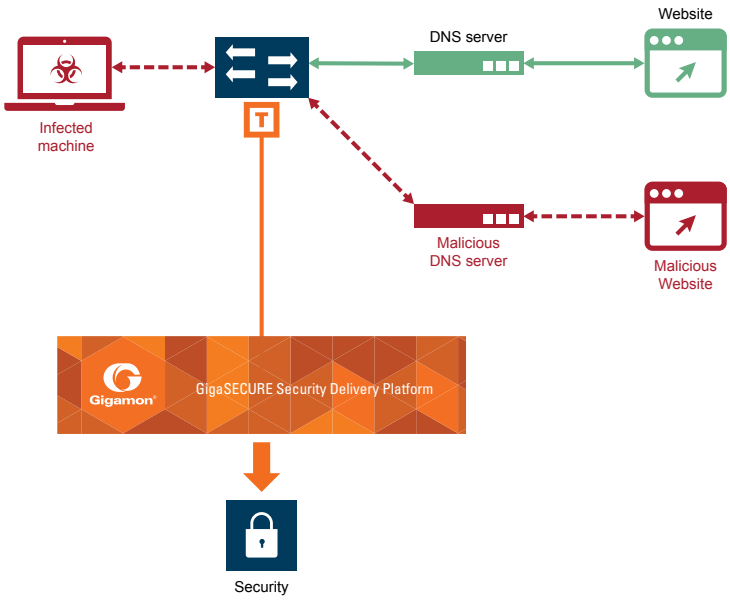


Figure 6: Identifying hijack attacks

Use Case Seven (DNS):

Detect Tunnels and Prevent Bypass of Security Controls

DNS tunnels provide an easy way to bypass security controls. Even networks that require authentication, like Wi-Fi hotspots, can carry traffic tunneled in DNS packets because, generally, DNS is always permitted. Configuring IPFIX to send DNS information to a SIEM or collector would indicate the number of transmission requests and enable an analyst or software engineer to quickly identify a tunnel.

In most cases, organizations need to collect DNS logs and parse them for an unusually large number of domains from a single IP, the host tunneling traffic. But what if they can't see those logs or don't have access to them? With GigaSECURE, they can flip a switch, capture that traffic right off the wire, and send it to a SIEM or even a script that checks for the presence of a user tunneling traffic.

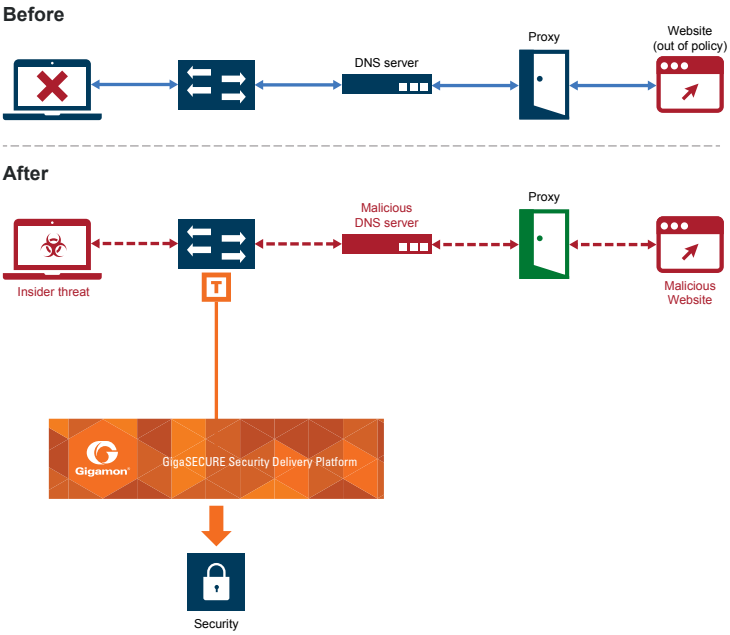


Figure 7: Identifying DNS tunneling

Use Case Eight (HTTP):

Use HTTP Return Codes to Detect Malware on the Network

If malware is on a network, it may be sending connect requests to domains or URLs that are not yet active or have ceased to be active.

GigaSECURE captures HTTP return codes. With them, it can see if there is a high number of 404 (Not Found error message) or 403 (Forbidden HTTP) return codes, which could indicate that a bot, rather than a human, is making the requests. Humans won't request pages that are not present more than a couple of times before giving up and moving on.

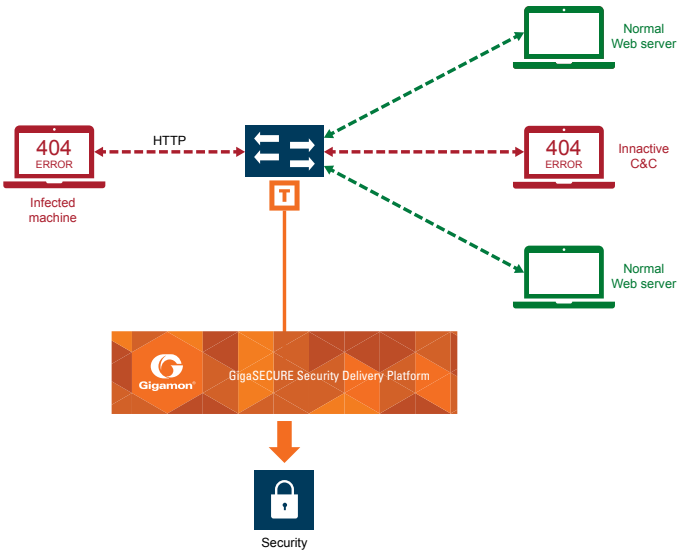


Figure 8: Detecting malware via HTTP return codes

Use Case Nine (SSL):

Uncover Malware Hidden in Encrypted Traffic Using SSL Certificate Information

According to the *Dell 2016 Annual Threat Report*, nearly 65 percent of Internet traffic is now encrypted. While encryption is good when it protects organization's confidentiality and data, it is not so good when it hides malware. And studies also show that SSL is one of the fastest growing attack vectors. In fact, Gartner's report, *Security Leaders Must Address Threats from Rising SSL Traffic*, says that, by 2017, more than 50 percent of network attacks will use encrypted traffic to bypass controls.

SSL certificates can be signed by Certificate Authorities (CA) and contain information about the server. In order to get a CA to sign a certificate, the administrator must first create a keypair with a Certificate Signing Request (CSR). This CSR is sent to CA and, for a fee, is signed. The purpose is to verify the identity of the server and help prevent incidents like DNS hijacking attacks.

SSL certificates can also be self-signed. This means that encrypted conversations can occur, but the identity of the server cannot be verified. These "clown suit certs" are used by administrators who are building new systems to test, but are not concerned with verifying the server's identity.

The presence of an SSL certificate on its own might be a clear indicator of issues, often serious ones. Once traffic is encrypted, tools designed to detect inconsistent behavior within the data are rendered useless.

However, in order for the encryption process to begin, the two ends must negotiate what encryption to use and part of the negotiation process involves exchanging SSL certificates. If you are not expecting self-signed certificates, SSL connections to IP addresses, or SSL connections on non-standard ports, monitoring for SSL with GigaSECURE would alert your tools of a possible serious issue.

In any case, an organization can glean a tremendous amount of valuable information from certificates, including whether or not the certificate is signed by a CA or self-signed, the length of the key, the strength of the algorithm, the associated domain, and the certificate's validity start and end date. For example, when running a website, it's important for organizations to know—and be notified in advance—when a certificate is expiring. When a certification expires and becomes invalid, website users will receive error messages.

Organizations might also want to know if there are weak algorithms in use and if there are servers in need of patching. In some cases, an unpatched Web server might accept a message from a client to force the use of old encryption for that session. This misconfigured server issue would be spotted when the cyphers were negotiated.

By monitoring SSL certificate exchanges, GigaSECURE can detect encrypted traffic on non-standard ports and, because it sees unsampled IPFIX, it can send a notification if a certificate is exchanged on any port, and uncover potential issues.

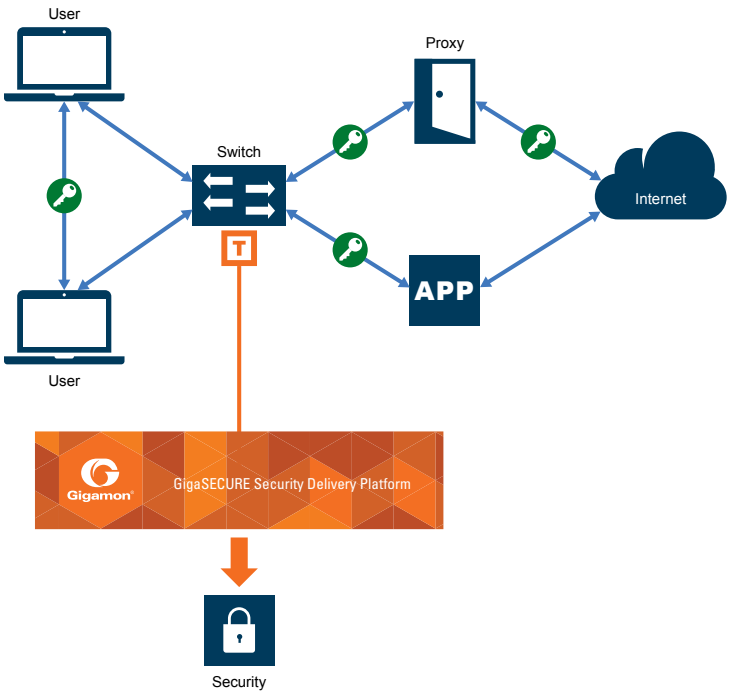


Figure 9: Monitoring SSL certificate exchanges to uncover hidden malware

Summary

In short, an organization's ability to grab information off the wire—and not be limited to sending log files—can dramatically improve its security posture. And with the GigaSECURE® Security Delivery Platform, users can see 100 percent of every traffic packet sent and received as well as deliver relevant, summarized information to SIEMs and collectors.

The GigaSECURE IPFIX-generation capability is designed to address stringent security use cases. Not only does GigaSECURE generate IPFIX records by looking at all packets on the wire, but it also provides customizable templates for IPFIX records. This allows information security professionals to accurately fingerprint network activity and provide better detection of command and control activity, as well as enable rapid forensics and anomaly detection.

About the Gigamon Metadata Engine

Gigamon's GigaSECURE® is a Security Delivery Platform (SDP) that provides pervasive reach across the infrastructure, spanning cloud, on-premise data centers/locations, and remote sites in order to bring the network to the security and performance management devices that require access to it in order to conduct their functions. Security appliances specifically, simply connect into the GigaSECURE platform to receive a high-fidelity stream of relevant traffic from anywhere in the network infrastructure. One key pillar of the GigaSECURE Security Delivery Platform is the ability to generate summary information about packets from network traffic and put it in the IPFIX format. This functionality provides security and behavioral analytics products with valuable unsampled information about traffic without impacting the performance of infrastructure. Recently, Gigamon transformed IPFIX generation into a full-blown Metadata Engine that sits inside the GigaSECURE SDP and serves as the single source of network truth for all kinds of information about applications, users, and devices, and sends relevant information to the security tools connected to the SDP.

The GigaSECURE Metadata Engine provides powerful infrastructure by which to collect network metadata elements (inclusive of IPFIX) that are critical for security operations. It offers a single source of valuable data for both network and security operations.

About Gigamon

Gigamon provides active visibility into physical and virtual network traffic enabling stronger security and superior performance. Gigamon's Visibility Fabric™ and GigaSECURE®, the industry's first Security Delivery Platform, deliver advanced intelligence so that security, network and application performance management solutions in enterprise, government and service provider networks operate more efficiently and effectively. With over ten years' experience, Gigamon solutions are deployed globally across vertical markets including over seventy-five percent of the Fortune 100.

For more information visit **www.gigamon.com**

3206-02 08/16