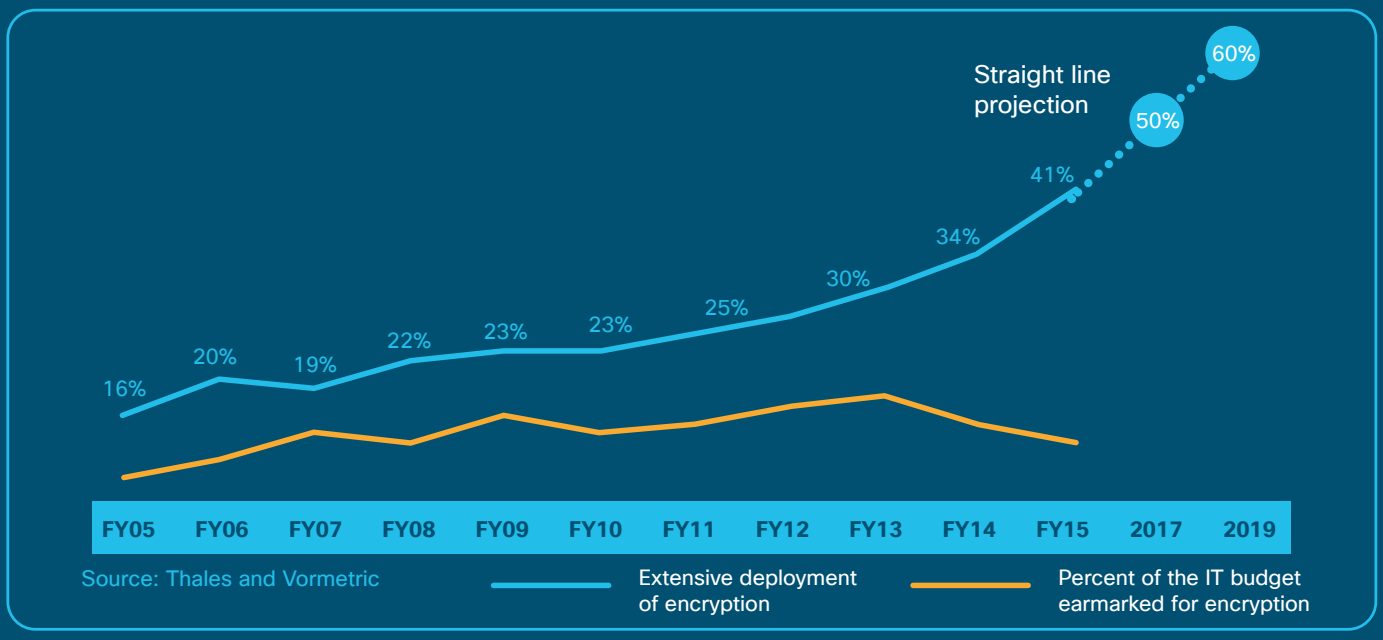
Encrypted Traffic Analytics

Introduction

The rapid rise in encrypted traffic is changing the threat landscape. As more businesses become digital, a significant number of services and applications are using encryption as the primary method of securing information. More specifically, encrypted traffic has increased by more than 90 percent year over year, with more than 40 percent of websites encrypting traffic in 2016 versus 21 percent in 2015. Gartner predicts that by 2019, 80 percent of web traffic will be encrypted.

Encryption technology has enabled much greater privacy and security for enterprises that use the Internet to communicate and transact business online. Mobile, cloud and web applications rely on well-implemented encryption mechanisms, using keys and certificates to ensure security and trust. However, businesses are not the only ones to benefit from encryption. Threat actors have leveraged these same benefits to evade detection and to secure their malicious activities. Figure 1 shows the economic impact of such attacks.

Encryption is changing the threat landscape



Contents

Visibility across the network is getting increasingly difficult and our traditional means of detection cannot assume that data is available for inspection. We need to be able to simultaneously assess how much of our digital business is protected and unprotected by encryption while also assessing what traffic is malicious and what is benign.

Gartner believes that half of malware campaigns in 2019 will use some type of encryption to conceal malware delivery, command and control activity, or data exfiltration.

Figure 1. Economic impact of malicious attacks

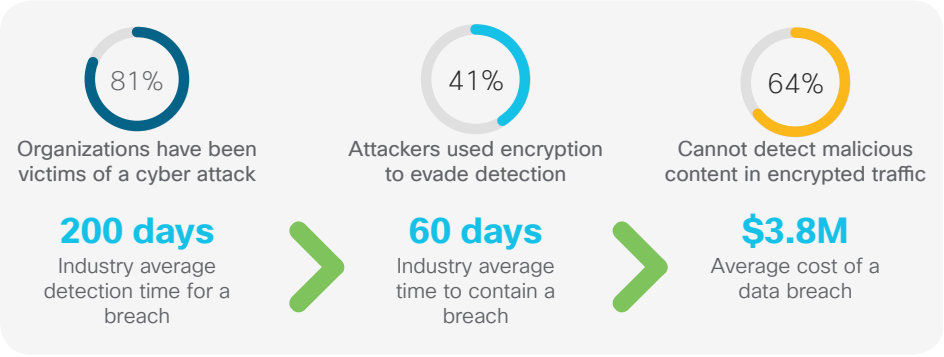


Table 1 describes the new threat vectors that are based on the nature of encrypted traffic.

Table 1. New threat vectors based on nature of encrypted traffic

Uninspected Encrypted Traffic	Threats
Employees' web browsing over HTTPS	▪ Malware infection ▪ Covert channel with the command and control server ▪ Data exfiltration
Employees on an internal network connecting securely to network edge (DMZ) servers	Lateral expansion from infected hosts
Internet users connecting to the enterprise's public servers using encrypted protocols	Reduced defense-in-depth, with only one protection technology inspecting incoming traffic

Challenges of encrypted traffic security

The majority of organizations today do not have a solution to detect malicious content in encrypted traffic. They lack the security tools and resources to implement a solution that can be deployed throughout their network infrastructure without slowing down the network.

Traditional threat inspection with bulk decryption, analysis and reencryption is not always practical or feasible, for performance and resource reasons. In many cases, however, advanced analytic techniques can be used to identify malicious flows for further inspection using decryption techniques.

On any given day, no one knows how much of their digital business is in the clear versus encrypted. If traffic is encrypted, the encryption is typically done to meet compliance requirements that mandate specific security policies.

Overview of Encrypted Traffic Analytics

Traditional flow monitoring provides a high-level view of network communications by reporting the addresses, ports and byte and packet counts of a flow. In addition, intraflow metadata, or information about events that occur inside of a flow, can be collected, stored and analyzed within a flow monitoring framework. This data is especially valuable when traffic is encrypted, because deep-packet inspection is no longer viable. This intraflow metadata, called Encrypted Traffic Analytics, is derived by using new types of data elements or telemetry that are independent of protocol details, such as the lengths and arrival times of messages within a flow. These data elements have the attractive property of applying equally well to both encrypted and unencrypted flows.

Using these data elements or intraflow telemetry to identify malware communication in encrypted traffic means Encrypted Traffic Analytics can maintain the integrity of the encrypted flow without the need for bulk decryption (Figure 2). Table 2 lists the benefits of using Encrypted Traffic Analytics.

Figure 2. Encrypted Traffic Analytics – technical solution overview

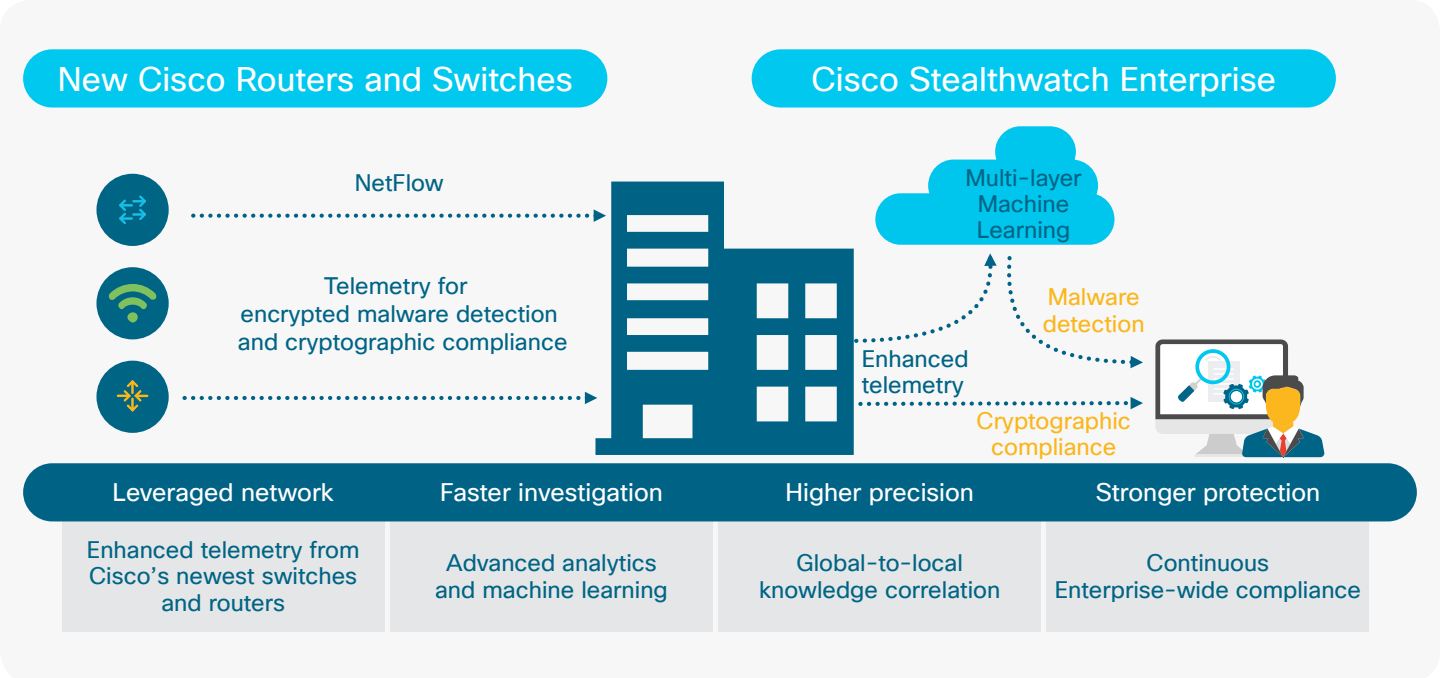


Table 2. Benefits of using Encrypted Traffic Analytics

Benefits
▪ Security visibility: Gain insight into threats in encrypted traffic using network analytics. Obtain contextual threat intelligence with real-time analysis correlated with user and device information. ▪ Cryptographic assessment: Ensure enterprise compliance with cryptographic protocols and visibility into and knowledge of what is being encrypted and what is not being encrypted on your network. ▪ Faster time to response: Quickly contain infected devices and users. ▪ Time and cost savings: Use the network as the foundation for the security posture, capitalizing on security investments in the network.

Encrypted Traffic Analytics – New data elements for encrypted traffic

Encrypted Traffic Analytics focuses on identifying malware communications in encrypted traffic through passive monitoring, the extraction of relevant data elements and supervised machine learning with cloud-based global visibility.

Transport Layer Security (TLS) is a cryptographic protocol that provides privacy for applications. TLS is usually implemented on top of common protocols such as HTTP for web browsing or Simple Mail Transfer Protocol (SMTP) for email. HTTPS is the use of TLS over HTTP. This is the most popular way of securing communication between a web server and client and is supported by most major web servers.

Encrypted Traffic Analytics extracts four main data elements: the sequence of packet lengths and times, the byte distribution, TLS-specific features and the initial data packet. Cisco's unique Application-Specific Integrated Circuit (ASIC) architecture provides the ability to extract these data elements without slowing down the data network.

- Sequence of Packet Lengths and Times (SPLT): SPLT conveys the length (number of bytes) of each packet's application payload for the first several packets of a flow, along with the interarrival times of those packets. SPLT can be represented as an array of packet sizes (in bytes) along with an array of times (in ms) representing the time since the previous packet was observed.
- Byte distribution: The byte distribution represents the probability that a specific byte value appears in the payload of a packet within a flow. The byte distribution of a flow can be calculated using an array of counters. The major data types associated with byte distribution are full byte distribution, byte entropy and the mean/standard deviation of the bytes. For example, using one counter per byte value, an HTTP GET request, "HTTP/1.1.", can be calculated by incrementing the corresponding counter once for the "H," then incrementing another counter twice for the two consecutive "T" s and so on. Although the byte distribution is maintained as an array of counters, it can easily be turned into a proper distribution by normalizing by the total number of bytes.
- Initial Data Packet (IDP): IDP is used to obtain packet data from the first packet of a flow. It allows extraction of interesting data such as an HTTP URL, DNS hostname/address and other data elements. The TLS handshake is composed of several messages that contain interesting, unencrypted metadata used to extract data elements such as cipher suites, TLS versions and the client's public key length.

Appendix A shows a detailed table of new data elements.

Encrypted Traffic Analytics - Components

Enhanced NetFlow

In the NetFlow architecture, data is transmitted from exporter to collector in sets of records. Each record in a data set has the same format, which is specified by its template. The data record consists of a series of NetFlow information elements or "fields," and a specific ID value is assigned to each field. The ID values for information elements may be globally defined and archived by the Internet Assigned Numbers Authority (IANA), or they may be enterprise specific and defined by individual organizations.

NetFlow templates use several globally defined elements administered by IANA. Some of the global elements, such as IP addresses and Layer 4 port numbers, form a familiar 5-tuple that is used as a unique flow identifier (flow key). Additional elements are used to report basic packet/octet statistics and timestamps.

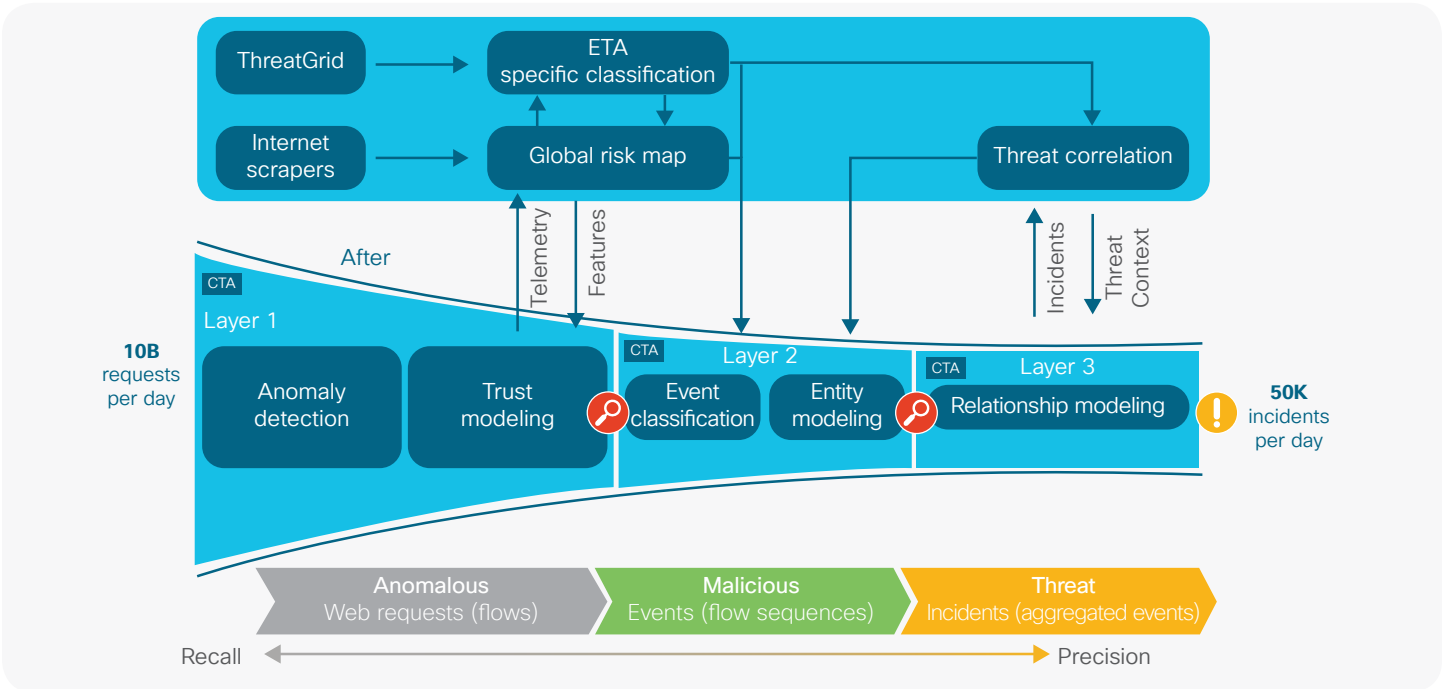
These globally defined elements are enhanced with vendor-specific (Cisco vendor ID) data elements described earlier and in Appendix A. The vendor-specific data elements provide insights into threats and vulnerabilities in encrypted traffic using Cisco Stealthwatch Enterprise®.

Stealthwatch Enterprise

Cisco Stealthwatch Enterprise uses NetFlow, proxy servers, endpoint telemetry, policy and access engines, traffic segmentation and more to establish baseline "normal" behavior for hosts and users across the enterprise. Stealthwatch can correlate traffic with global threat behaviors to automatically identify infected hosts, command and control communication and suspicious traffic.

Stealthwatch maintains a global risk map – a very broad behavioral profile about servers on the Internet, identifying servers that are related to attacks, may be exploited, or may be used as a part of an attack in the future (Figure 3). This is not a blacklist, but a holistic picture from a security perspective. Stealthwatch analyzes the new encrypted traffic data elements in enhanced NetFlow by applying machine learning and statistical modeling. The global risk map and Encrypted Traffic Analytics data elements reinforce using advance security analytics. Rather than decrypting the traffic, Stealthwatch uses machine learning algorithms to pinpoint malicious patterns in encrypted traffic to help identify threats and improve incident response.

Figure 3. Stealtwatch Enterprise Multi-layer Machine Learning



The Security Insight dashboard on the Stealthwatch Management Console (SMC) provides a view of affected users identified by risk type. An expanded dashboard provides detailed information regarding the top risk escalations and relative threat exposure. Table 3 lists some high-risk threats that use encrypted command and control communications.

Figure 4. Stealthwatch security insight dashboard

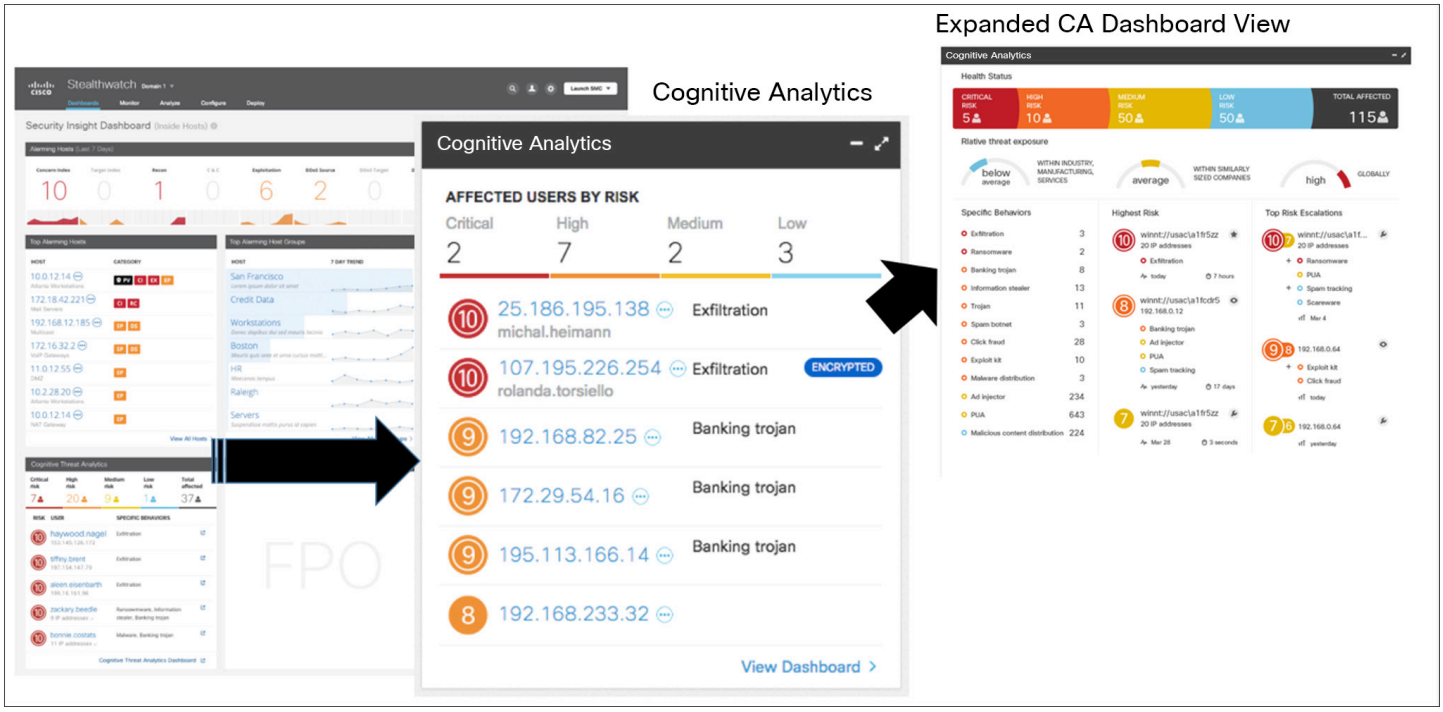


Table 3. Examples of high-risk threats using encrypted command and control

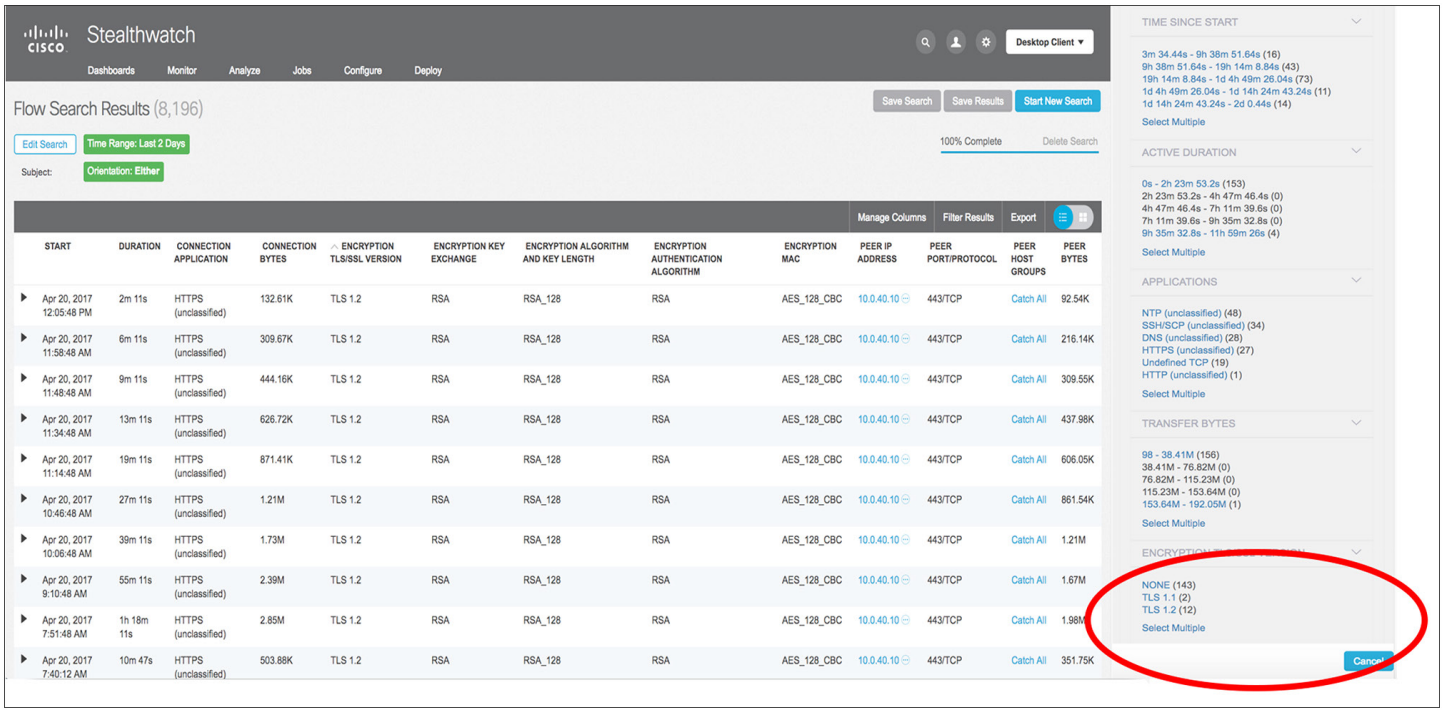
Name	Type
Gamarue/Andromeda	Modular botnet
Sality	File infecter, modular botnet
Necurs	Information stealer, backdoor, botnet
Rerdom	Click-fraud, botnet

Upon discovery, a malicious encrypted flow can be blocked or quarantined by Stealthwatch. Policy-driven remediation actions via pxGrid using Cisco Identity Services Engine (ISE) with Cisco TrustSec® and Software-Defined Access (SD-Access) simplify and accelerate network security operations.

Cryptographic assessment

Encrypted Traffic Analytics also identifies encryption quality instantly from every network conversation providing the visibility to ensure enterprise compliance with cryptographic protocols. It delivers the knowledge of what is being encrypted and what is not being encrypted on your network so you can confidently claim that your digital business is protected. This cryptographic assessment is displayed in Stealthwatch and can be exported via APIs to third-party tools for monitoring and auditing of encryption compliance (Figure 5).

Figure 5. Cryptographic assessment



Feature support

Cisco's newest networking equipment, starting with Cisco IOS® XE 16.6, will support an enhanced NetFlow with Encrypted Traffic Analytics capability.

1. Compatible Cisco equipment supporting enhanced NetFlow with Encrypted Traffic Analytics:

- Switches: Cisco Catalyst® 9300 Series (starting with the Cisco IOS XE 16.6.1 release) and the 9400 Series (starting with the Cisco IOS XE 16.6.2 release)
- Routers: ASR 1001-X, ASR 1002-X, ASR 1001-HX, ASR 1002-HX, ASR1000 RP2, ASR1000 RP3, ASR1000 ESP-40, 4221 ISR,4321 ISR, 4331 ISR, 4351 ISR, 4431 ISR, 4451-X ISR, ISR 1000 series routers, Integrated Services Virtual Router (ISRv) including the 5000 Enterprise Network Compute System, Cloud Services Router (CSR) 1000V (starting with the Cisco IOS XE 16.6.2 release)

2. Stealthwatch gains additional machine learning and statistical modeling capabilities (in release 6.9.2) to analyze enhanced NetFlow with Encrypted Traffic Analytics.

Conclusion

In summary, the network is now an even more advanced security sensor, capable of detecting threats in encrypted traffic. A Cisco Digital Network Architecture-ready infrastructure turns the network into an end-to-end sensor and enforcer that detects, contains and prevents emerging, sophisticated security threats.

Appendix A

Data Elements Extracted by Encrypted Traffic Analytics.

Data Element Name	Description
Sequence of Packet Lengths and Times (SPLT)	An array of LENGTH values followed by an array of INTERARRIVAL TIME values describing the first N packets of a flow that carry application payload. Each LENGTH is encoded as a 16-bit integer to form a 20-byte array. Immediately following this, each INTERARRIVAL TIME is encoded as a 16-bit integer to form another 20-byte array.
Byte distribution	A histogram giving the frequency of occurrence for each byte value or (range of values) in the first N bytes of application payload for a flow. Each "frequency of occurrence" is represented as a 16-bit integer.
Initial Data Packet (IDP)	The content of the first packet of this flow that contains actual payload data, starting at the beginning of the IP header.
TLS records	An array of LENGTH values, followed by an array of INTERARRIVAL TIME values, followed by an array of CONTENT TYPE values, followed by an array of HANDSHAKE TYPE values. These arrays describe the first N records of a TLS flow.
TLS record lengths	A sequence of record lengths for up to the first N records of a TLS flow.
TLS record times	A sequence of TLS interarrival times for up to the first N records of a TLS flow.
TLS content types	A sequence of ContentType values for up to the first N records of a TLS flow.
TLS handshake types	A sequence of HandshakeType values for up to the first N records of a TLS flow.
TLS cipher suites	A list of up to N cipher suites offered by the client, or selected by the server in a TLS flow.

Data Element Name	Description
TLS extensions	An array of LENGTH values followed by an array of EXTENSION TYPE values describing the TLS extensions observed in the Hello message for a TLS flow.
TLS extension lengths	A list of extension lengths for up to the first N TLS extensions observed in the TLS Hello message for a flow.
TLS extension types	A list of extension types for up to the first N TLS extensions observed in the TLS Hello message for a flow.
TLS version	The TLS version number observed in the TLS Hello message for a flow.
TLS key length	The length of the client key observed in the TLS ClientKeyExchange message.
TLS session ID	The session ID value observed (if any) in the TLS Hello message for a flow.
TLS random	The random value observed in the TLS Hello message for this flow.

References

1. Gartner: Security Leaders Must Address Threats from Rising SSL Traffic
2. Ponemon Institute: Uncovering Hidden Threats Within Encrypted Traffic, 2016
3. NSS Labs: TLS/SSL: Where Are We Today? The Encrypted Web Part 1 – An Upward Trajectory
4. Identifying Encrypted Malware Traffic with Contextual Flow Data, Blake Anderson and David McGrew, AISEC '16