Subscriber Data Correlation

Application of Cisco Stealthwatch to Service Provider mobility environment

Introduction

With the prevalence of smart mobile devices and the increase of application usage, Service Providers (SPs) are facing burdens on their Long-Term Evolution (LTE) networks. By 2020, more than 70 percent of Internet traffic will come from smart devices, of which more than 40 percent will be using LTE data.

In addition, Law Enforcement Agencies (LEAs) conduct electronic surveillance using processes for analysis and investigations. Legislation and regulations are being enforced that require SPs and ISPs to implement electronic surveillance, and in many occasions, SPs are forced to pay huge fines for not being compliant, leading to business losses.

This white paper describes the solution provided by Cisco® technologies to help SPs comply with LEA mandates and avoid losses due to penalties enforced by regulations. This document is based on a solution that has been tested in production to analyze an SP's mobile subscriber data traffic and help achieve regulatory compliance.

The main business drivers and problems that are addressed by this solution are:

- **Visibility of mobile data:** In most cases, SPs face a lack of visibility into traffic that is generated from subscriber's smart devices connected to their network. This traffic can be the result of normal day-to-day activities as well as suspicious activity.
- **Compliance:** Governing entities are imposing policies forcing SPs to keep logs on Internet traffic coming from mobile subscribers. These logs are often required to be kept and stored using external systems.
- **Reduce business cost and risk:** SPs are subject to substantial fines if they are not able to comply with regulations and laws. These fines can be in the range of millions of U.S. dollars, leading to severe financial loss.

Contents

Solution requirements

To comply with the regulations, the solution should provide end-to-end visibility on traffic flows originating from subscriber devices and going to the public Internet, in addition to:

- Correlating events in a human readable format
- Using a search function to simplify tracking of data flow connection
- Showing the public/internal IP address association, Gateway GPRS (General Packet Radio Service) Support Node (GGSN), Mobile Station International Subscriber Directory Number (MSISDN), and URL visited as well as timestamps in a single window In addition to the above, the solution should be capable of correlating data from multiple sources and able to scale to a service provider network considering these technical requirements and challenges:
- **Complexity of the network:** The network has multiple entry and exit points, and any solution must be flexible enough to cater to the network design. This includes multiple locations, complicated routing, and other factors.
- **Multiple data sources:** The information required for building a complete picture resides in several repositories. For example, user authentication data is in a different server than the network policy. These various data sources need to be combined to provide context-aware visibility.
- **NAT:** More and more service providers are implementing Network Address Translation (NAT) on their Gi interface (the interface between the Internet and the SP's network). This has multiple benefits in securing the network and conserving public IP address space. However, this limits the visibility. As such, any solution must be able to correlate flows for before and after the NAT interface.
- **Data retention and export:** Flow data collected by the solution have to be stored for months, depending on local laws and company policy. The system provides a capability to provide end-to-end visibility and correlation of the collected information and reduce the volume of the information that needs to be exported and stored on other systems for long periods.

Solution

Solution brief

This solution correlates data flows to provide SPs with visibility into the mobile traffic passing through their network and the capabilities to attribute traffic using a public IP address to the originating subscriber device.

The main objective of the solution is to correlate the flow information at the Gi interface with the internal user identity. The two major components that need to be interoperated are:

- **Internal Authentication, Authorization, and Accounting (AAA):** This component contains the authentication data for internal subscribers. The data we are most interested in is the MSISDN, internal IP Address, and the domain the user is authenticated to.

- **NAT gateway:** This is where the flows will be directed for NAT as well as security policies. This gateway will export the flow information to the solution.

If more granular information is required, including the URLs accessed as well as latency information, a sensor can be deployed to generate flow data from raw traffic sent to it through a Switched Port Analyzer (SPAN). This information can then be correlated with the identity data from the AAA server as well as the flow information from the NAT gateway.

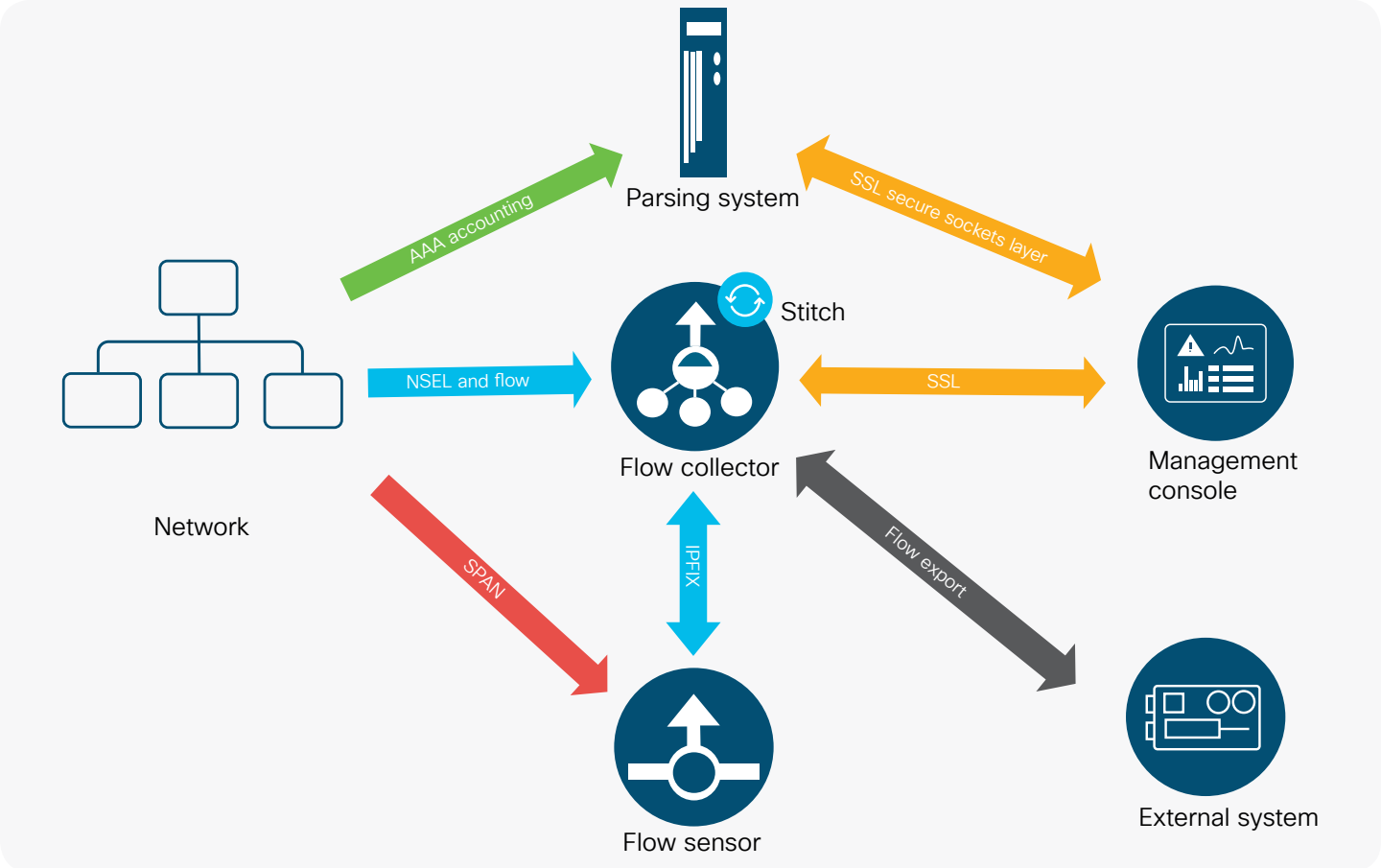
Data flow

The solution requires the correlation of multiple data sources containing multiple information elements into a single conversational record using Cisco Stealthwatch®, an industry-leading network visibility and security analytics technology:

- MSISDN information exported from RADIUS accounting systems
- Network traffic flow information that is collected using NetFlow
- NAT information gathered from NetFlow Security Event Logging (NSEL)
- Destination URL information detected through deep packet inspection on network spanned traffic

Below is a flow diagram explaining how the data is being transferred and correlated:

Figure 1. Data flow



Flow diagram details:

1. The AAA accounting information is forwarded to the parsing system that will extract the following fields:
   - MSISDN
   - Internal IP address
   - GGSN
2. The parsing system will then forward this information to the Cisco Stealthwatch Management Console (SMC) using logs.
3. The SMC will correlate this data and send it to the Cisco Stealthwatch Flow Collector for stitching with the network flow information.
4. The Flow Collector is collecting NSEL/NetFlow information from the firewalls. NSEL records contain the pre- and post-NAT IP addresses as well as traffic flow information, including:
   a. Outside global address (post NAT)
   b. Inside address (pre NAT)
   c. Date and time of the flow
   d. Source port
   e. Destination IP address and port
   f. Protocol
   g. Device interface
   h. Time stamp
   i. Bytes exchanged
5. SPAN information taken from the network will let the Cisco Stealthwatch Flow Sensor perform deep packet inspection and extract HTTP URL information from the traffic when a mobile device connects to the Internet. The Cisco Stealthwatch Flow Sensor will then create an IP Flow Information Export (IPFIX) record, add the extracted URL information to it, and export it to the Flow Collector for stitching.
6. The Cisco Stealthwatch Flow Collector will perform the most important task in the data flow and correlation process by deduplicating and stitching information received from all the below sources:
   a. Pre- and post-NAT and flow information received from the Adaptive Security Appliance (ASA)
   b. MSISDN and GGSN information received from the SMC
   c. URL and flow information received from the Flow Sensor
7. Finally, the data stitched will be stored on the Flow Collector. The SMC will be able to perform queries on the Flow Collector database and display the information in a human-readable format and in a single window. In addition, the Flow Collector will be able to export the flow information to an external system, which in this case was used to retain the data over a long period of time.

Technologies

- **NetFlow**: This is a built-in feature in most network infrastructure devices that provides network traffic metadata. Switches, routers, and firewalls support this feature and provide valuable information about traffic, including:
  - Source IP and port
  - Destination IP and port
  - Protocol
  - Device interface
  - Time stamp
  - Bytes exchanged
- **ASA/Fire Power Threat Defense (FTD) NSEL:** The ASA/FTD implementation of NSEL is a stateful IP flow-tracking method that exports records that indicate significant events in a flow. In stateful flow tracking, tracked flows go through a series of state changes. NSEL events are used to export data about flow status and are triggered by the event that caused the state change. The significant events that are tracked include flow-create, flow-teardown, and flow-denied.

  The ASA implementation of NSEL provides the following major functions:
  - Keeps track of flow-create, flow-teardown, and flow-denied events and generates appropriate NSEL data records.
  - Defines and exports templates that describe the progression of a flow. Templates describe the format of the data records that are exported through NetFlow. Each event has several record formats or templates associated with it.
  - Tracks configured NSEL collectors and deliver templates and data records to these configured NSEL collectors through NetFlow.
  - Uses NetFlow version 9 to export the NSEL data and includes the NAT translation information from the ASA firewall.

- **Flow Sensor DPI:** Flow Sensor uses Deep Packet Inspection (DPI) techniques to examine the packets of the traffic that is sent to it using SPAN. The sensor will examine the data and the headers in order to detect and gather application layer information in addition to network performance–related metrics to calculate network round trips and server response times. In addition, the DPI performed on the HTTP traffic will help the sensor extract HTTP traffic responses and URLs and web domain formation.
- **Accounting data processing and export:** Accounting data is processed to retrieve contextual information contained in some accounting and authentication protocols (RADIUS). The contextual information can provide additional details such as authenticated subscriber identity (MSISDN), authentication point (GGSN), and assigned internal address (private IP). This data is then exported in a log or flow format to be consumed and stitched into the traffic by the solution.
- **Flow export:** The flow export capability is a feature that is available in Cisco Stealthwatch version 6.8.2 that allows the Flow Collector to export the stitched and correlated flows to another system for various type of use, such as specific security or network analysis or even extending the data retention period beyond the Flow Collector limits.

Solution details

This section provides details on the components and architecture implemented for the solution.

Components

Cisco Firepower 9300

The Cisco Firepower® 9300 Security Appliance is a security service platform that enables multiple security functions to exist in a single physical appliance. It is designed with the SP in mind to allow for high scalability and clustering, which further increases the throughput available. Cisco Firepower 9300 runs the Firepower eXtensible Operating System (FXOS) as the base operating system and runs the ASA application in its security modules to perform the network policy, NATand flow export functions required for this solution.

ASA/FTD application

The ASA or FTD application is the stateful firewall function of Cisco Firepower 9300. The ASA or FTD has been enhanced to serve several SP functions such as CGNAT, routing, stateful inspection, security gateway, and more.

AAA RADIUS proxy

The AAA performs the authentication of the subscribers and keeps a record of these authentications. The accounting information is needed for the solution to correlate the subscriber to the flows. There are several possible methods for transmitting this information. For this solution, the simplest and most effective method was to use the AAA RADIUS proxy function to send a copy of the accounting records to the log parsing system.

Flow sensor (added separately to Cisco Stealthwatch)

Flow Sensor is a physical appliance–based solution that connects passively to the network through a SPAN, mirror, The traffic received is analyzed using a combination of Deep Packet Inspection (DPI) and behavioral analysis to identify applications and protocols in use, regardless if the traffic is in plain text or uses advanced encryption and obfuscation techniques. It also extracts HTTP URL information from HTTP traffic, in addition to gathering packet-level performance statistics to detect network round trips and server response times for a client-server conversation. The extracted information is then exportable in a flow format such as IPFIX.

The parsing system

The parsing system is a virtual system that runs a Linux distribution and is capable of consuming multiple type of logs, flows, and accounting information and transforms it into multiple types of outputs such as flows and logs. The main use case of such a system is to format outputs from a system "A" to another type of output that can be consumed by another system "B."

SMC (included in core Cisco Stealthwatch product)

The SMC is a physical appliance or a virtual system that represents the centralized management, configuration, and reporting system for Stealthwatch.

The SMC provides advanced build-it views and graphics to help understand network and security problems detected by Cisco Stealthwatch, in addition to the ability to build customized queries and reports on the data stored in the system. Additionally, advanced drill-down capabilities are available and help to view traffic information down to a single flow.

Flow collector (included in core Cisco Stealthwatch product)

Flow Collector collects and analyses multiple types of flows (NetFlow, IPFIX, sFlow, etc.) from existing network infrastructure to provide a complete picture of everything happening in an environment. The Flow collector has multiple features and capabilities such as:

- **Telemetry stitching and deduplication:** This capability transforms the unidirectional flows collected from multiple network devices that are related to the same traffic flow into a bidirectional flow conversation that contains statistical traffic information on a conversation between a client and server. This feature reduces flow information false positives and allows efficient flow storage.
- **NAT stitching:** If the traffic is passing through a NAT device, flow information collected from network devices before and after the NAT operation will not have the same source and destination IP addresses. Some NetFlow and NSEL NAT-capable devices such as the Cisco ASA can export pre- and post-NAT information in the flow record in addition to security decisions done on the flow (flow permitted or denied). Flow Collector will stich the pre- and post-NAT IP addresses and security information and retains the bidirectional flow information from source to destination in a single record.
- **Visibility and security:** Flow Collector baselines, categorizes, and analyzes flows at the network and host level. Based on algorithms, Flow Collector will detect anomalies in an environment from a network or security perspective.
- **Storage:** Flow Collector stores formatted and stitched flows to its local disks and provide the capability to search for this data easily for forensics purposes. In this solution, Flow Collector is used to store the flows for a short-range use, and the flows are exported to another system to provide longer storage capability.

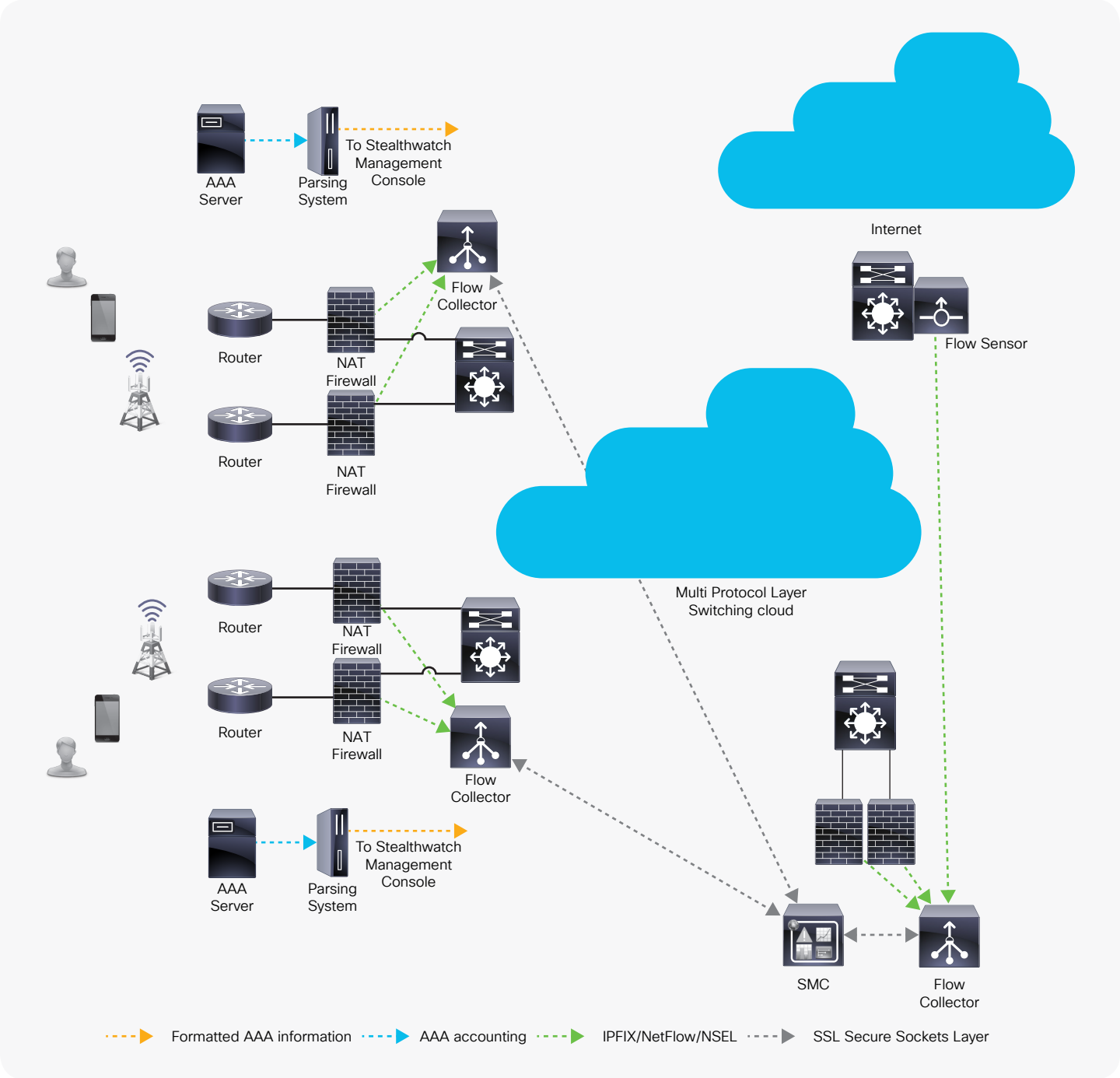
External storage system

Flow collector storage provides the capability to search for the information simply and quickly. In addition, the export capability allows the solution to export the flows to an external storage system to provide long-period storage depending on the customer and regulation requirements. In the solution described here, an external third-party storage system was used to consume the exported flows by the flow collectors and store them for extended storage.

Architecture

The diagram below shows the architecture that was used for this solution:

Figure 2. Solution design



The solution was implemented in an SP network where the Multiprotocol Label Switching network is connecting multiple locations to a main data center and to the Internet services node.

Each site is located in a main area and serves subscribers that are authenticated with the GGSN through an AAA RADIUS system. The RADIUS system is forwarding the accounting information using the RADIUS proxy technique to the parsing system. The parsing system is consuming the AAA information and transforms it to a format that can be processed by Cisco Stealthwatch. In addition, ASA firewalls, in each location, will send NSEL and flow information to a Flow Collector that performs the stitching, analysis, and storage functions on flows and on the AAA information received from the parsing system through the SMC.

The SMC performs the centralized configuration and management role of the distributed Flow Collectors and is sharing the AAA information gathered from each site with all the Flow Collectors. In addition, the SMC provides a single dashboard to query all the collectors and displays a unified view of all the flows collected from all the distributed locations.

At the Internet serving node, Flow Sensors use the mirrored traffic to perform DPI, extract the URL and network performance parameters, and then send it in IETF IPFIX format to the Flow Collector to stitch this additional contextual and application layer information into the flow record. Flow Sensors are only located at the Internet edge node because all the traffic that is connecting to the Internet is established through this node.

Conclusion

In this execution model, Cisco Stealthwatch is designed to provide telemetry-based end-to-end visibility into mobile data traffic. Multiple data types form different sources are correlated to provide traffic details, subscriber info, and application context for every traffic flow passing through the network. The output result is displayed and can be queried through an easy-to-use user interface. The correlated metadata is exportable from the system for long-term storage and is easily accessible for analysis and forensics purposes.

Cisco Stealthwatch helps SPs achieve compliance with regulations and conform to LEA requirements related to electronic surveillance by delivering the following capabilities:

- **Visibility of mobile data:** Cisco Stealthwatch will provide end-to-end visibility into the traffic generated from subscribers using the SP's infrastructure.
- **Compliance:** Cisco Stealthwatch empowers SPs to comply with policies enforced by governing entities and laws by providing the capability to traffic metadata records for subscribers connected to the SP's network.
- **Reduced business cost and risk:** This solution reduces business risks and cost by reducing the likelihood of fines and penalties with minimal investment, which helps SPs to continue achieving regular business growth and expand their mobile business.