CHAPTER 1

Overview

This chapter describes the FlowCollector application, which is used with the NetFlow services data export feature on Cisco routers and Catalyst 5000 and 6000 series switches. This chapter includes information on the following topics:

What Are NetFlow Services?

NetFlow services consist of high-performance IP switching features that capture a rich set of traffic statistics exported from routers and switches while they perform their switching functions. The exported NetFlow data consists of traffic flows, which are unidirectional sequences of packets between a particular source device and destination device that share the same protocol and transport-layer information. The captured traffic statistics can be used for a wide variety of purposes, such as network analysis and planning, network management, accounting, billing, and data mining.

Because of their unidirectional nature, flows from a client to a server are differentiated from flows from the server to the client. Flows are also differentiated on the basis of protocol. For example, Hypertext Transfer Protocol (HTTP) Web packets from a particular source host to a particular destination host constitute a separate flow from File Transfer Protocol (FTP) file transfer packets between the same pair of hosts.

Routers and switches identify flows by looking for the following fields within IP packets:

- Source IP address
- Destination IP address
- Source port number
- Destination port number
- Protocol type
- Type of service (ToS)
- Input interface.

Catalyst 5000 series switches can identify flows by looking at a subset of these fields. For example, they can identify flows by source and destination address only.

Note For Catalyst 5000 series switches, the analog to NetFlow services is integrated Multilayer Switching (MLS) management. Included are products, utilities, and partner applications designed to gather flow statistics, export the statistics, and collect and perform data reduction on the exported statistics. MLS management then forwards them to consumer applications for traffic monitoring, planning, and accounting.

NetFlow Services Device and IOS Release Support

NetFlow functionality is currently available with the following Cisco devices:

- Cisco routers
- Catalyst 5000 series switches equipped with a NetFlow feature card (NFFC)
- Catalyst 6000 series switches

You can find the most up-to-date information available to help you determine the compatibility among different Cisco hardware platforms, Cisco IOS software releases, and supported NetFlow data export versions at the following URL:

http://www.cisco.com/warp/public/732/Tech/matrices/netflow_matrix.shtml

Note Except for descriptions requiring references to specific router or switch platforms, the remainder of this chapter and the remaining chapters of this guide use the term export device instead of the terms router and switch.

NetFlow Data Export

NetFlow data export makes NetFlow traffic statistics available for purposes of network planning, billing, and so on. An export device configured for NetFlow data export maintains a flow cache used to capture flow-based traffic statistics. Traffic statistics for each active flow are maintained in the cache and are incremented when packets within each flow are switched. Periodically, summary traffic statistics for all expired flows are exported from the export device by means of User Datagram Protocol (UDP) datagrams, which FlowCollector receives and processes.

How and When Flow Statistics Are Exported

NetFlow data exported from the export device contains NetFlow statistics for the flow cache entries that have expired since the last export. Flow cache entries expire and are flushed from the cache when one of the following conditions occurs:

- The transport protocol indicates that the connection is completed (TCP FIN) plus a small delay to allow for the completion of the FIN acknowledgment handshaking.
- Traffic inactivity exceeds 15 seconds.

For flows that remain continuously active, flow cache entries currently expire every 30 minutes to ensure periodic reporting of active flows.

NetFlow data export packets are sent to a user-specified destination, such as the workstation running FlowCollector, either when the number of recently expired flows reaches a predetermined maximum, or every second-whichever occurs first. For a Version 1 datagram, up to 24 flows can be sent in a single UDP datagram of approximately 1200 bytes. For a Version 5 datagram, up to 30 flows can be sent in a single UDP datagram of approximately 1500 bytes. For a Version 7 datagram, up to 27 flows can be sent in a single UDP datagram of approximately 1500 bytes. For a Version 8 datagram, the number of flows sent in a single UDP datagram varies by aggregation scheme. See Appendix B, "NetFlow Export Datagram Formats," for details on all versions of the NetFlow data export format.

NetFlow Export Data Format

NetFlow exports flow information in UDP datagrams in one of four formats: Version 1 (V1), Version 5 (V5), Version 7 (V7), or Version 8 (V8).

The Version 1 (V1) format is the original format supported in the initial NetFlow releases. The Version 5 (V5) format is an enhancement that adds Border Gateway Protocol (BGP) autonomous system information and flow sequence numbers. The Version 7 (V7) format is an enhancement that exclusively supports Cisco Catalyst 5000 series switches equipped with a NetFlow feature card (NFFC). V7 is not compatible with Cisco routers. The Version 8 (V8) format is an enhancement that adds router-based aggregation schemes. Versions 2, 3, 4, and 6 are not supported by FlowCollector. For more information on the distinctions among the four format types, see Appendix B, "NetFlow Export Datagram Formats."

The following types of information are part of the detailed traffic statistics:

- Source and destination IP addresses
- Next hop address
- Input and output interface numbers
- Number of packets in the flow
- Total bytes (octets) in the flow
- First and last time stamps of packets that were switched as part of this flow
- Source and destination port numbers
- Protocol
- Type of service (ToS)
- Source and destination autonomous system (AS) numbers, either origin or peer (present in V5 and select V8 datagrams)
- Source and destination prefix mask bits (present in V5, V7, and V8 datagrams)
- Shortcut router IP address (present in V7 on Cisco Catalyst 5000 series switches only).

Caution Throughout this publication there are numerous examples of FlowCollector input commands and output results. Included are examples of IP addresses. Be aware that IP address examples are not usable IP addresses. The examples do not represent real-life configurations.

What Is FlowCollector?

FlowCollector provides fast, scalable, and economical data collection from multiple export devices exporting NetFlow data records. Figure 1-1 shows an example of a typical NetFlow data export scheme. In it, various export devices send export data to user-specified FlowCollector UDP ports.

Figure 1-1 FlowCollector Overview

Each of the export devices in this example is configured for NetFlow data export. Part of the configuration information for each export device includes the IP address and the UDP port number (a logical port designator) that identify FlowCollector as the receiver of flows from this export device. The UDP port number is a user-configurable designator: you can configure FlowCollector to listen for flows on a number of different UDP ports, and then configure your export devices so that each device exports flows to a dedicated UDP port, or have a number of devices export flows to the same, shared UDP port.

After you configure and start FlowCollector, it listens to the user-specified UDP ports for exported flows from the export devices you have configured for NetFlow data export.

FlowCollector performs the following functions:

- NetFlow data collection from multiple export devices
- Reduction in data volume through filtering and aggregation
- Hierarchical data storage (helps client applications retrieve data)
- File system space management

FlowCollector collects and summarizes (aggregates) data into data files based on user-defined criteria specified in a FlowCollector thread. A thread is an aggregation task defined by a set of user-configurable attributes that specify how FlowCollector aggregates the traffic flows stored on the workstation. Two key thread attributes are:

- Aggregation schemes-Define how to summarize the traffic flows stored on the workstation
- Filters-Define the information that is accepted or rejected by a thread

FlowCollector provides a set of predefined aggregation schemes to help you collect NetFlow export data and summarize the data (that is, aggregate the flows). You can choose one or more of these aggregation schemes to customize FlowCollector for your operating context.

You can also use filters with aggregation schemes to include or exclude certain types of NetFlow data. FlowCollector provides a set of predefined filters to provide further help in refining the range and type of traffic statistics collected and summarized. You can also define your own filters to customize FlowCollector.

For more information about threads, aggregation schemes, and filters, see Chapter 5, “Customizing FlowCollector.”

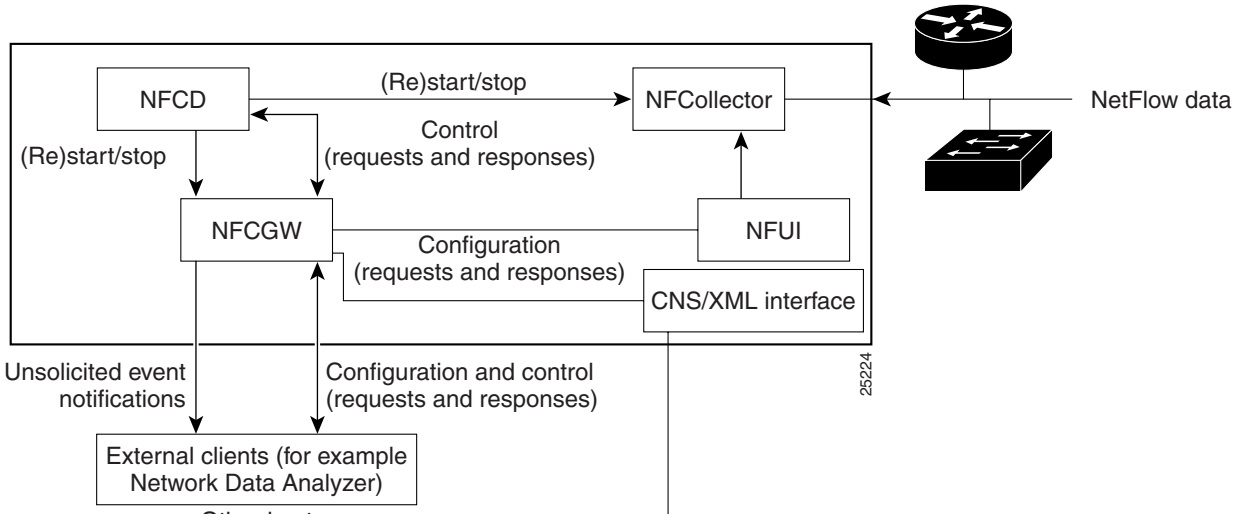
FlowCollector Architectural Overview

FlowCollector consists of four subsystems:

- Collector (NFCollector)
- Gateway (NFCGW)
- Daemon (NFCD)
- User Interface (NFUI)

These subsystems work together to provide FlowCollector functionality, including data collection, the user interface, configuration and control, and so forth. They also provide a communications link with the Network Data Analyzer (NDA) application, and custom client applications developed to interface with FlowCollector. See Figure 1-2 for a graphical representation of the FlowCollector system architecture, and see Appendix C, "FlowCollector Configuration and Control Protocol" for information on the FlowCollector development protocol.

Figure 1-2 FlowCollector System Architecture



Collector

The Collector (NFCollector) is the heart of the system. This subsystem collects NetFlow data, aggregates (or summarizes) that data, and filters specified data from supported Cisco routers and switches. The data collected by NFCollector is stored in data files that are organized in an easy-to-use directory structure.

Gateway

The Gateway (NFCGW) is a subsystem that interfaces with external client applications that *talk* to NFCollector. External client applications include the NDA application and other custom applications developed using FlowCollector configuration and control protocol language. See Appendix C, "FlowCollector Configuration and Control Protocol" for complete details on the protocol language. NFCGW is socket-based, and it accepts requests from client applications (like NDA) to query or change NFCollector configuration parameters and also broadcasts unsolicited event notifications (UENs) using UDP datagrams.

Daemon

The Daemon (NFCD) monitors the operational status of both NFCollector and NFCGW. NFCD is user-configurable. It provides the ability to start, and if necessary, restart NFCollector and NFCGW when the systems shut down because of operational problems. If NFCollector or NFCGW should terminate for any reason, NFCD restarts the terminated processes. NFCD is installed as a Daemon on the FlowCollector workstation and is customizable through the **nfcd.config** configuration file. See the "nfcd.config" section on page 5-3 for details on the **nfcd.config** file.

User Interface

The User Interface (NFUI) is used to query NFCollector for runtime statistics and to perform configuration tasks. See Chapter 3, "Using the FlowCollector User Interface," for complete details on the user interface.