IdeaData Pty Limited

Technical Specification

NetFlow Auditor Data Collection System

NetFlow Auditor Support
2008

Contents

Overview 4
Description 5
NetFlow Auditor Real-Time 6
Network Forensics and Security 6
Key Features 6
NetFlow Auditor Long-Term 7
Network Auditing and Trending 7
Key Features 7
Key Benefits 8
Scalability 8
Data Compression 8
Data Value-Add 8
Information Redundancy 8
SQL Data Access 8
Multiple Operating Systems Support 8
Implementation 9
Passive Monitoring 9
End-to-End Automation 9
Plug-ins 9
Security & Privacy 9
Granularity of Reporting 10
Technology 11
Data Sources 11
NetFlow 11
Tools 11
Development 12
MySQL 12
Apache Tomcat 12
NetFlow Auditor 13
Collectors 13

Router Plug-ins 13
Back End 14
A Scalable Approach 14
A Flexible Approach 14
A Secure Approach 14
A Unique Approach 15
A Smart Approach 15
Aggregation 16
Front End 17
1 - Zooming In 17
2 - Drilling Down 19
Part 3 – Traffic Filtering 20
System Requirements 22
Operating System 22
Database Engine 22
Web Server 22
Client 22
Hardware 22

Overview

Information technology (IT) infrastructure in today's complex environments necessitates the need to identify, capture and report on every conversation across the entire network.

Considering the facts, there is a growing need for network administrators to be able to report on traffic traversing all sites. This requires information collection from every router, switch, and the recording of every conversation, both large and small; NetFlow Auditor does all of this, now!

NetFlow Auditor's methods of collection, aggregation, and storage of network data are advantageous for the following reasons – speed (requires minimal computing), size (requires minimal data-storage), granularity (requires minimal data-storage and speed of access to granular data is fast) and privacy (if the data is stored in raw format NetFlow Auditor would be unable to protect the privacy of users if required).

NetFlow Auditor offers customers an access management solution which enables corporations to efficiently and accurately track, trace, monitor, alert and manage traffic traveling inside their network; and outside to the Internet. Helping various parts of the business achieve visibility to solve technical issues and to address management with accurate information about business systems to support infrastructure decisions and risk.

- IT infrastructure expenditure is soaring at an exponential rate to meet the increased, and sometimes unsubstantiated, demands of its users and applications. We all face the challenge of achieving an efficient, high bandwidth technology infrastructure. However, the corporate dollar spend on establishing a competitive IT environment will put further pressures on balance sheets as data usage continues to grow.

- With increased diversity and volume of network traffic, be it for voice, video, server based or peer to peer; it is essential to understand who is talking to who; who is talking to where; and how much talking is going on inside your network. Measurability, accountability and knowledge of a company's network use or abuse has until now been at best on an "ad hoc" basis.

- Organizations continue to store confidential and sensitive information on their networks without knowing who is accessing this information.

- Organizations are relying more and more on a stable and secure IT network for everyday business continuity and need to identify external threats from Viruses, Trojans, Spyware, DoS and hacking.

Description

The NetFlow Auditor suite of solutions utilizes NetFlow, Ethernet or Proxy data from today's networks to generate a unique database. This database is used to store point to point information on all network sessions flowing through the network over an extended period of time. With this information, NetFlow Auditor can present valuable network usage reports that are suitable for a wide variety of audiences.

By utilizing NetFlow Auditor, it is now possible to provide complete network coverage from a data source that has already been installed and paid for – the network routing hardware itself.

NetFlow Auditor makes every network conversation accessible to a wide range of technical and non-technical audiences and integrates with any existing application that can be made to call a URL.

NetFlow Auditor Real-Time

Network Forensics and Security

NetFlow Auditor Real-Time provides enhanced troubleshooting and forensic capabilities for today's complex networks. NetFlow Auditor Real-Time is also an intuitive and intelligent agent that learns and builds a baseline of the traffic flows occurring on any network and can alert network management on bursts, scans and peer-to-peer (P2P) traffic. The level of granularity and the ability to drill into the traffic is unprecedented with every flow being stored for every minute for up to one month[1].

NetFlow Auditor Real-Time maintains baselines with recent entries and traffic patterns and will use previously collected records or input from a Long-Term system to build a new baseline immediately based on ultra long-term network auditing. It will start building a baseline from the first hour it has been implemented with ability to build alert unique data specific triggers that will ensure that the longer NetFlow Auditor Real-Time has been collecting the smarter it becomes.

NetFlow Auditor Real-Time enables real-time dynamic drill down to the root causes of network traffic and automatically produces alerts and reports that can aid in troubleshooting network problems and determine the source of spurious traffic. NetFlow Auditor Real-Time lets you drill down to see all traffic and lets you see the most noisy hosts and traffic consumers on your network. You can drill down to see any aspect of the data applications used and source and destination addresses contacted with full understanding of the traffic impact.

Key Features

- Real-time visibility of every network flow recorded, per minute.
- Historical visibility of every network flow recorded, per minute (standard 7 days, typical 4 weeks).
- Full integration with long term historical visibility of every network flow recorded, per hour (standard 12 months, typical 3-4 years).[2]
- Secure integration with any third party management product, using URLs
- Ability to filter reports based on fields stored in NetFlow/Ethernet/Proxy data.
- Fully customized reports scheduled to email, file location or Intranet portal; hourly, daily, monthly, quarterly or yearly in both CSV and html formats.
- No changes to existing network infrastructure required.
- No impact to network performance.

[1] Standard data storage us configured for seven days.
[2] Available when installed with DigiToll NetMeasure.

NetFlow Auditor Long-Term

Network Auditing and Trending

NetFlow Auditor Long-Term is a robust fully automated end-to-end Network knowledge base to enable you to plan and profile the use of any element within your network. NetFlow Auditor Long-Term equips management with the knowledge to better understand IT resource and network usage enabling an organization to maximize operational efficiencies.

NetFlow Auditor Long-Term takes the guess work out of trending with the ability to retain granular information of the network and key business systems usage over a long term enabling your organization to effectively measure usage, trending patterns, flows, packets, baselines, averages, peaks and troughs, and standard deviations. Having access to long-term historical data ensures that an organization can forecast and plan for their future resource requirements of key business systems.

NetFlow Auditor Long-Term elegantly captures; stores and produces granular traffic from numerous data sources including Cisco NetFlow, Ethernet and Proxy Logs and stores this into a database for long-term analysis. Our patented[3] method for recording a data transfer substantially reduces the space required to store and process the data. Other software requires large computing resources that can quickly become overwhelmed by the size of the data and the accuracy and timeliness of reporting.

Today's digitally orientated smart businesses are realizing the need to track data usage for capacity planning and long-term forensics and to feel and understand the dynamic pulse of their business. NetFlow Auditor Long-Term information is targeted towards key managers including the Capacity Planner, CEO, CFO and CIO.

Key Features

- Long term historical visibility of every network flow recorded, per hour (standard 12 months, typical 3-4 years).
- Secure integration with any third party management product, using URLs
- Ability to filter reports based on fields stored in NetFlow/Ethernet/Proxy data.
- Fully customized reports scheduled to email, file location or printer; hourly, daily, monthly, quarterly or yearly in both CSV and html formats.
- No changes to existing network infrastructure required.
- No impact to network performance.

[3] Patents Pending. PCT/AU2003/001418.

Key Benefits

Scalability

NetFlow Auditor is highly scalable in large data management and multi-collector / multi-database distribution. NetFlow Auditor Intelligent Agents reduce large data sets using a patented[3] method of describing and managing the transfer of data. The data is inserted directly into a database in a 'described' manner. NetFlow Auditor's scalability is a direct result of architecting intelligent distributed collector points to perform processing and aggregation. This allows for the ability to distribute tasks over multiple agents thus increasing the throughput of the overall system making NetFlow Auditor completely scalable.

Data Compression

NetFlow Auditor Intelligent Agents reduce large data sets using a patented[3] method of describing data. This means that even though the data is not stored in a raw form, all data can be accessed as if it were in raw format. In fact NetFlow Auditor retains greater granularity in the data and for a longer period than any existing product in the market place. NetFlow Auditor has complete flexibility in the degree of granularity of data collection actually desired at the specified collection point. After the conversion of data into NetFlow Auditor compressed format the original data can be automatically stored in a ZIP/GZIP file format to enable absolute backup opportunity. These files can be automatically decompressed and read back into a NetFlow Auditor collector for reprocessing if the network information needs to be accessed.

Data Value-Add

NetFlow Auditor has the ability to correlate and associate other data sources thereby adding new meaning, clarification and classification of the data. Data from a router can be correlated with data from an authentication device or data from publicly accessible Internet databases such as SNMP, DNS, QoS descriptions and other data sources.

Information Redundancy

NetFlow Auditor's highly flexible intelligent agents can be set up to replicate data and thus provide an extra level of redundancy at any point of the data transfer. Standard database replication also adds further levels of redundancy for networks where the requirement for visibility is of greatest importance.

SQL Data Access

NetFlow Auditor utilizes SQL databases for the long-term storage of the network traffic data. SQL enables NetFlow Auditor to become an open system. The database schema is freely available to licensed users who can then use this information to enhance their own custom applications that require utilizing the data collected by NetFlow Auditor.

Multiple Operating Systems Support

The NetFlow Auditor Processor is a Java application focused on being platform independent. NetFlow Auditor currently runs on several flavors of UNIX including Solaris, Linux, in addition to

Windows platforms offering flexibility for multi-vendor support. NetFlow Auditor collectors being run in a distributed manner can be run over a mixed environment where some processors are being run on Linux or UNIX whilst others are being run on Windows.

Implementation

As a direct result of Intelligent Agent Technology, NetFlow Auditor is extremely simple to install, configure and use. The power of NetFlow Auditor's auto detection coupled with the simplicity of NetFlow Auditor's ability to perform network departmentalization and cost allocation allows users to see rich data about their network that was previously unavailable to them all in a remarkably short period of processing. This minimizes integration, implementation and configuration costs. Minimal training is required.

NetFlow Auditor requires minimal customization and is highly user intuitive. The enhanced forensic web interface provides users with an interactive data mining tool that is also very user intuitive and logical in its method of network data analytics.

Passive Monitoring

NetFlow Auditor is non-intrusive in its method of monitoring. NetFlow eliminates the need for multiple probes on a single switch and reduces the need for multiple probes in a LAN. Ethernet sniffing enables the collection of data where NetFlow is difficult to get. SNMP is used only as a method of automatically enhancing the meaning of the data using information already in the router/switch.

End-to-End Automation

NetFlow Auditor has been designed to be a fully automated system that can be set up to require absolutely no user interaction with the collection or reporting process. NetFlow Auditor possesses a wide set of automated tools that can perform tasks ranging from retrieving log files through secure FTP server for processing, to auto-discovering new users and IP addresses, generating daily reports as well as sending these reports through e-mail or to a web server. Once NetFlow Auditor has been configured; users have the luxury of receiving total network visibility performed by a fully automated system.

Plug-ins

NetFlow Auditor is powerful in being able to harnesses the ability to collect simultaneously from a wide array of different network data sources.

Security & Privacy

NetFlow Auditor enables the organization to discover and to extract documented proof of breaches of security that occur real-time or those that occur over a prolonged period.

To prevent misuse of the data collected by NetFlow Auditor, any transfer of data between collectors is performed using an encrypted stream. This will prove to be especially useful when the data needs to be transferred through a public or unsecured network (such as the internet) to an upstream collector. The encryption algorithm can detect data that has been tampered with and also doubles up as a data integrity checksum as corrupted data will not be decrypted successfully.

Usernames that are collected can also be encrypted in order to protect user privacy.

Granularity of Reporting

The ability to easily identify data at a granular level is critical as it pinpoints network accountability and visibility to the recipient of that information. NetFlow Auditor data is highly compressed; granular and available in both real-time and long-term historical formats. Therefore NetFlow Auditor reporting and alerting produces results for multiple network recipients from CEO / CFO to CIO / CTO Network Architects and Engineers. As a result, a single NetFlow Auditor cost effective package can be used where normally 3 to 5 different, non-scalable and costly packages would be used to attain some immediate requirements. This enables different organizations to satisfy immediate requirements for accountability and visibility below the macro level and the ability to scale to future requirements not immediately in their focus.

Communication Service Providers (CSP) and Organizational Internal Chargeback require granular information because they are able to substantiate and quantify actual usage costs. In the case of larger Communication Service Providers networks current methods of cost analysis have shown high-levels of discrepancy in billing. Granularity achieved by NetFlow Auditor assists in quickly reducing billing disputes and verification of true communication bandwidth charges. Granularity of NetFlow Auditor data enables clear breakdowns between types of usage and enables CSP's to differentiate pricing based on time periods or server content. In the case of Internal Chargeback, NetFlow Auditor granularity enables billing to occur to the smallest unit on the network such as an individual user. The ability for NetFlow Auditor to bill at a granular level allows complete flexibility on the elements that are billable e.g. Network, IP address, Port, Department, User, Mac Address, QoS or a combination thereof.

Granular traffic over the long term provides complete network visibility and enables the end-user to identify spurious traffic and reasons for service level impacts. Erratic traffic can utilize valuable and costly bandwidth resources.

The ability to perform network forensics at a granular level enables the organization to discover breaches of security that occur real-time or those that occur over a prolonged period.

Technology

Data Sources

NetFlow

- Source and destination IP address
- Source and destination TCP/User Datagram Protocol (UDP) ports
- Source and destination AS numbers
- Quality of service fields (ToS and Diff Serv)
- Packet, byte and flow counts
- Input and output interface numbers
- Encapsulated protocol (TCP/UDP)
- Cost Centre/Account Identification

* Refer to Cisco NetFlow Device Support at the end of this article.

Tools

NetFlow Auditor tools are split into:

- “Collectors” which collect, compress, normalize, correlate and aggregate source data.
- “Back End” which aggregates, processes and inserts (or if required transmits data to other remote NetFlow Auditor database servers for insertion) into the database from various collectors.
- “Front End” which allows interactive observation of NetFlow Auditor database.
- “Aggregation” which allows flexible aggregation of Collectors or other Hierarchical Collectors.
- "Reporting" sub-system that allows unattended reporting and alerting. Historical data remains in the SQL database server and can be checked at any time by using the interactive front end or pushed to users or other systems currently via email, html, or CSV file output.

Development

NetFlow Auditor is developed around the following open source technologies.

MySQL

The MySQL database has become the world's most popular open source database because of its consistent fast performance, high reliability and ease of use. It's used in more than 8 million installations ranging from large corporations to specialized embedded applications on every continent in the world.

Not only is MySQL the world's most popular open source database, it's also become the database of choice for a new generation of applications built on the LAMP stack (Linux, Apache, MySQL, PHP / Perl / Python.) MySQL runs on more than 20 platforms including Linux, Windows, OS/X, HP-UX, AIX, Netware, giving you the kind of flexibility that puts you in control.

MySQL was chosen by IdeaData because it provides the necessary performance, stability and reliability that have made NetFlow Auditor the number one choice for some of the largest networks in the world.

Apache Tomcat

Apache Tomcat is the servlet container that is used in the official Reference Implementation for the Java Servlet and Java Server Pages technologies. The Java Servlet and Java Server Pages specifications are developed by Sun under the Java Community Process.

Apache Tomcat is developed in an open and participatory environment and released under the Apache Software License. Apache Tomcat is intended to be a collaboration of the best-of-breed developers from around the world.

Apache Tomcat powers numerous large-scale, mission-critical web applications across a diverse range of industries and organizations.

Apache Tomcat was chosen by IdeaData as a robust and proven methodology to drive the web based front end reporting and analysis tool. Apache Tomcat gives NetFlow Auditor the flexibility to enable the end user to view network traffic and network detail from the same application.

NetFlow Auditor

Collectors

The data collection system of NetFlow Auditor is based on a modular and platform independent server/plug-in system written using Sun's Java language.

NetFlow Auditor plug-ins are based on the same core NetFlow Auditor Technology and thus can coexist. Each plug-in is an independent collector agent.

Different plug-ins support different data source types. This model assures users that NetFlow Auditor is expandable as new plug-ins are required.

Router Plug-ins

These plug-in provide NetFlow Auditor with real time "live" data processing capabilities.

Cisco NetFlow

The NetFlow plug-in collects NetFlow streams (UDP packets) by listening on a specific network interface to specified ports. The information stored include source and destination IP numbers, source and destination port numbers, traffic protocol, Interface addresses, ASN source and Destination, flow, volume of traffic, time resolution down to the hour (for long-term data) and time resolution down to and time resolution down to 1 minute (depending on the setup of the NetFlow sender). Note: Although different vendors may abide by the NetFlow standards set by Cisco there are often differences in the timeliness of accounting records appearing in the NetFlow stream. For example some devices only report the traffic between a source and destination once the conversation has ended.

NetStream (supported)

sFlow (unsupported)

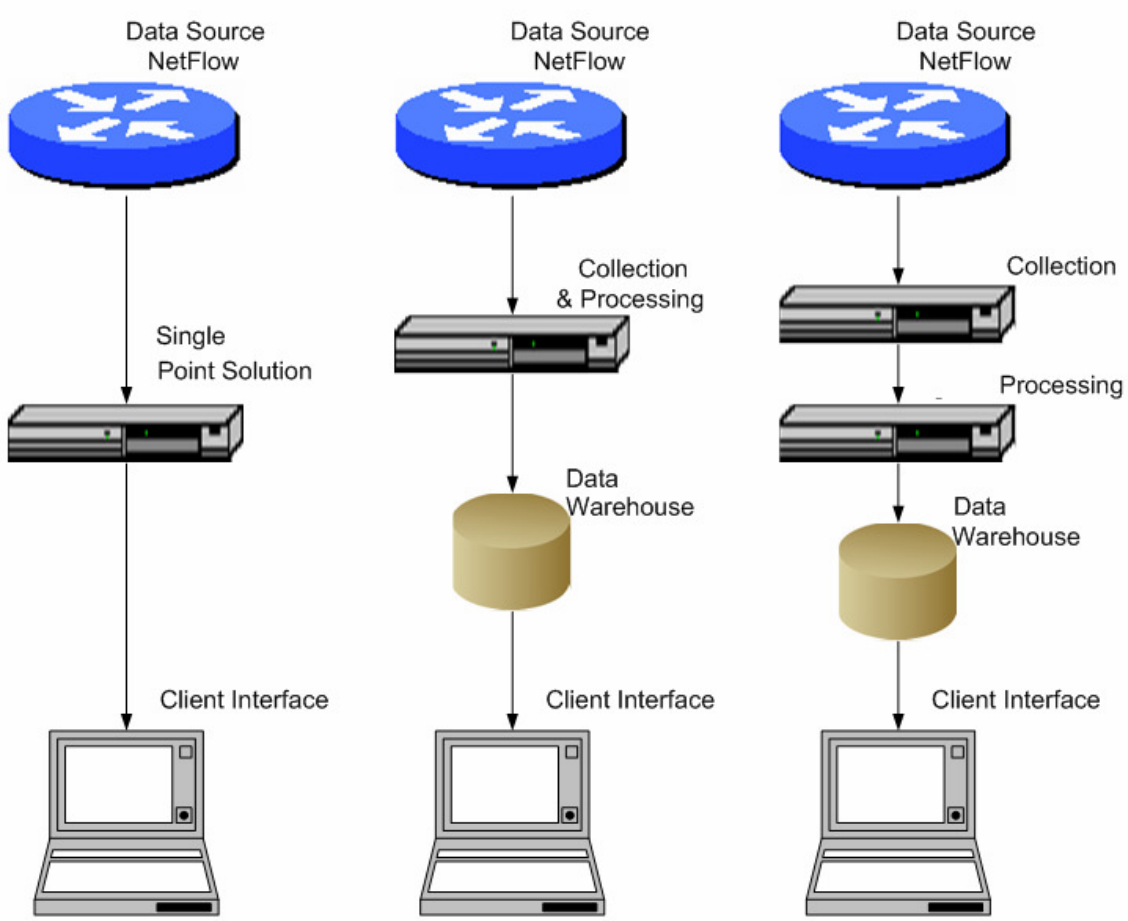
*IPFIX (supported)**

Back End

The NetFlow Auditor back end or 'engine', performs multiple distributed business critical, and network critical functions that include gathering, situation assessment, aggregation, alerting and report scheduling.

A Scalable Approach

The entire NetFlow Auditor engine will happily run on a single machine, however, for increased reliability and scalability, the NetFlow Auditor data collection and log file processing can be run on a separate machine to the database, which in turn may run on a separate machine to the "Front End". Thus CPU load can be distributed over several machines if required.



A Flexible Approach

NetFlow Auditor is the only network analytics tool of its type to give you the choice of NetFlow collection strategies. Collection occurs locally or distributed with High Availability Intelligent Agent Technology.

A Secure Approach

Transport of data from the collector to the database and between collectors is encrypted.

A Unique Approach

NetFlow Auditor can simultaneously collect information from different Network Gateways (i.e. Routers, Switches, etc...) and aggregates, processes and inserts the processed information into the slim line NetFlow Auditor SQL database. For some of these network gateways NetFlow Auditor performs "live" collection and for others NetFlow Auditor analyses historical logs. The NetFlow Auditor collection & processing methodology is a 3 step process.

A Smart Approach

Through IdeaData's patented5 live compression and normalization process NetFlow Auditor is able to keep data volume small, minimizing overhead on the WAN network links.

E.g. Raw NetFlow data covering a 500 user site = 500Meg/day, via NetFlow Auditor = 0.25-20Meg/day.

Data is "Normalized", compressed and encrypted before upload. The network data volume is kept small, minimizing WAN impact. Average ration of data stored is 0.1% of the original data volume. Transmission size and data capacity depends on volume of network traffic being monitored and tuning.

Long-Term would require approximately 200GB to store twelve months of data at full granularity in one hour increments for a 10,000 user environment.

Real-Time would require approximately 50GB to store seven days of data at full granularity in one minute increments for a 30,000 user environment.

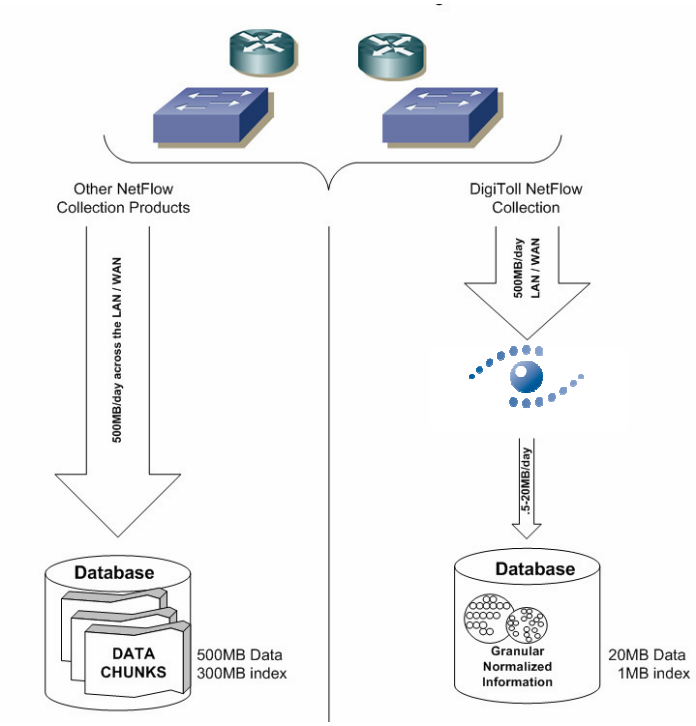
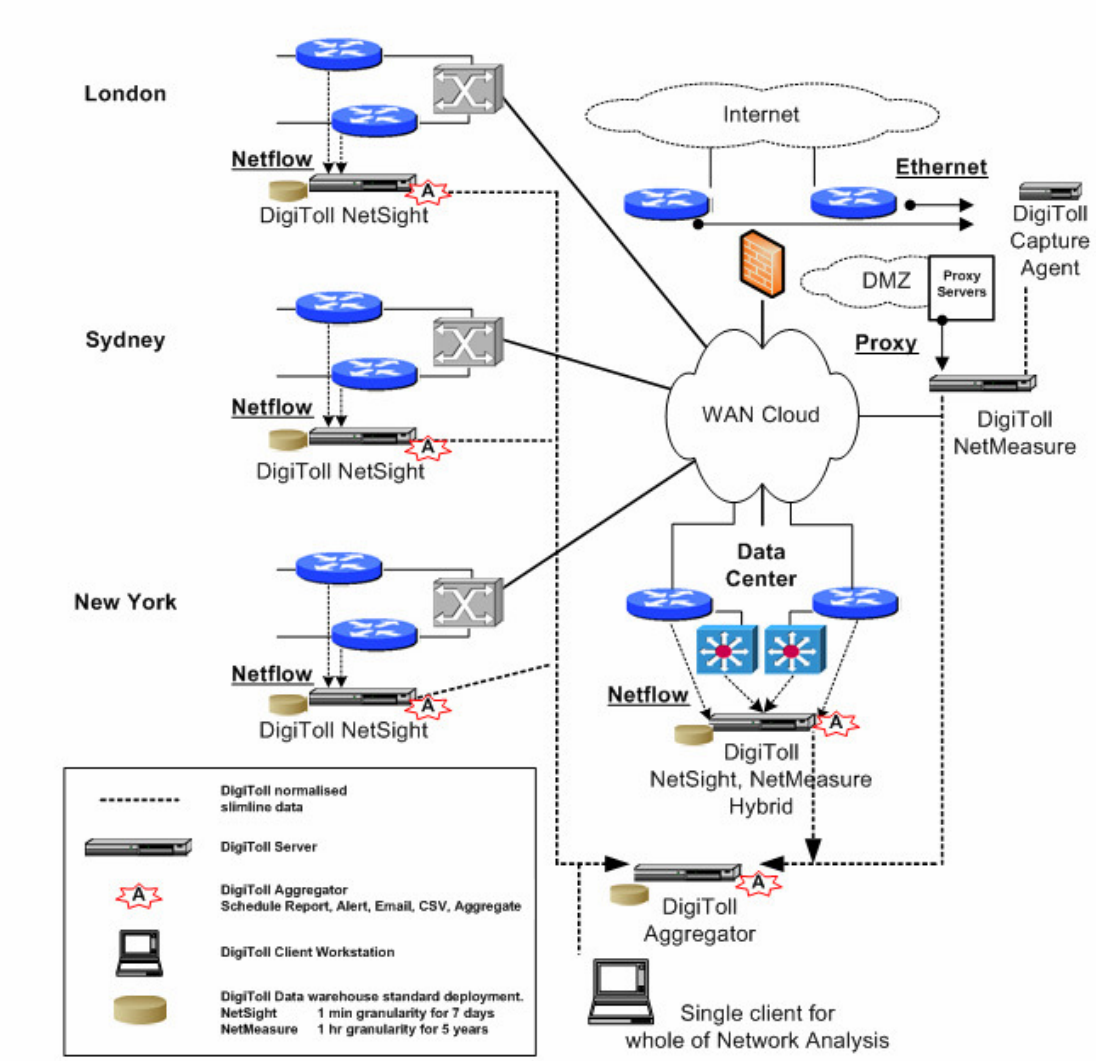


Figure 1: Other Cisco NetFlow Collection vs. NetFlow Auditor Patented Methodology

Aggregation

NetFlow Auditor Aggregation is used to scale NetFlow Auditor installations for large enterprise networks. This technology is used to aggregate NetFlow Auditor Real-Time or NetFlow Auditor Long-Term into a single data source giving a single NetFlow Auditor Long-Term view for your entire network.



Front End

Both NetFlow Auditor Real-Time and NetFlow Auditor Long-Term use a common single web enabled user interface. This interface is driven by Sun Java SE2 v1.5 or above.

The front end interface works in 3 basic methods

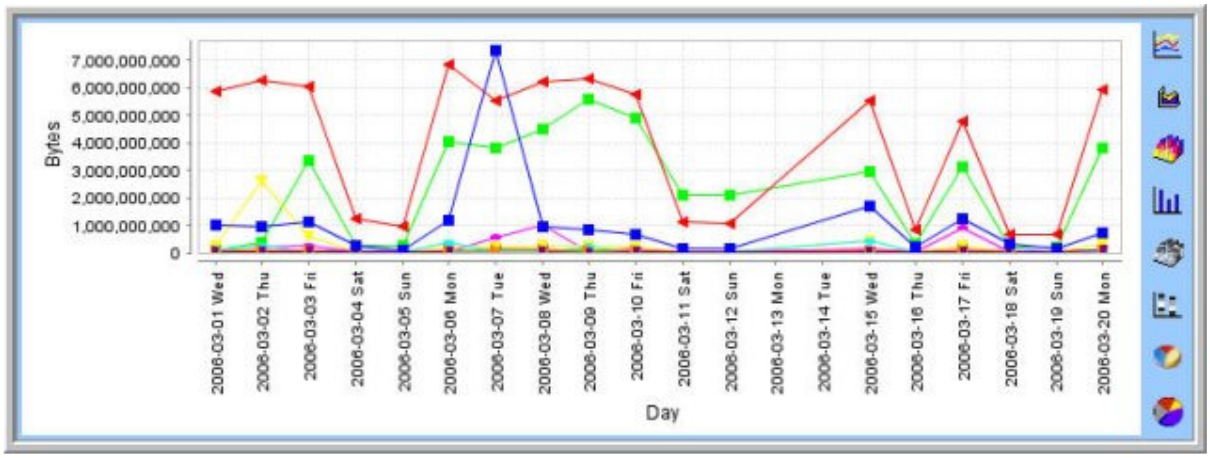
1 - Zooming In

The zooming functionality is available for the NetFlow Auditor charting applet. Users can zoom into all bar, line and area charts.

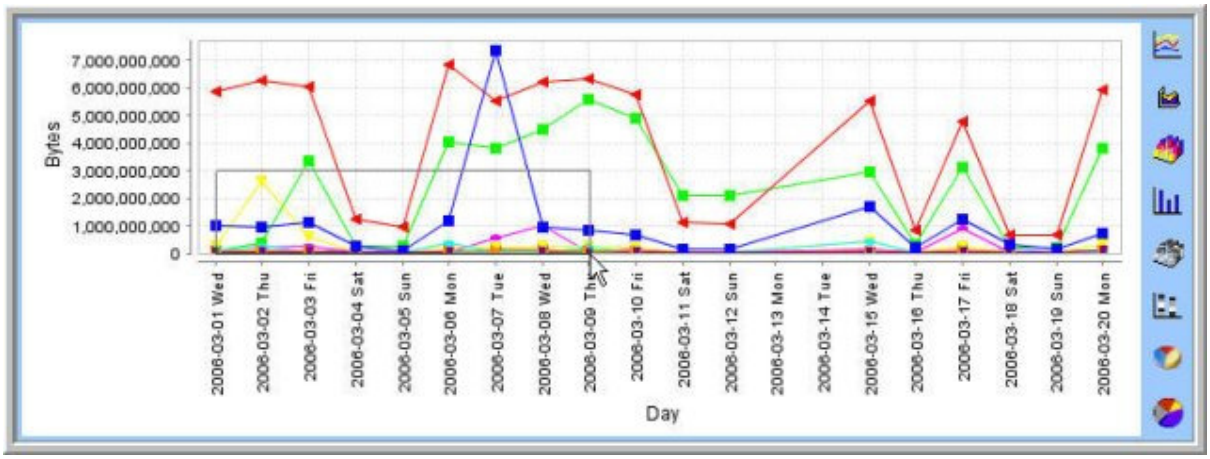
You can zoom into a report to reveal more intricate details.

To zoom down into a report:

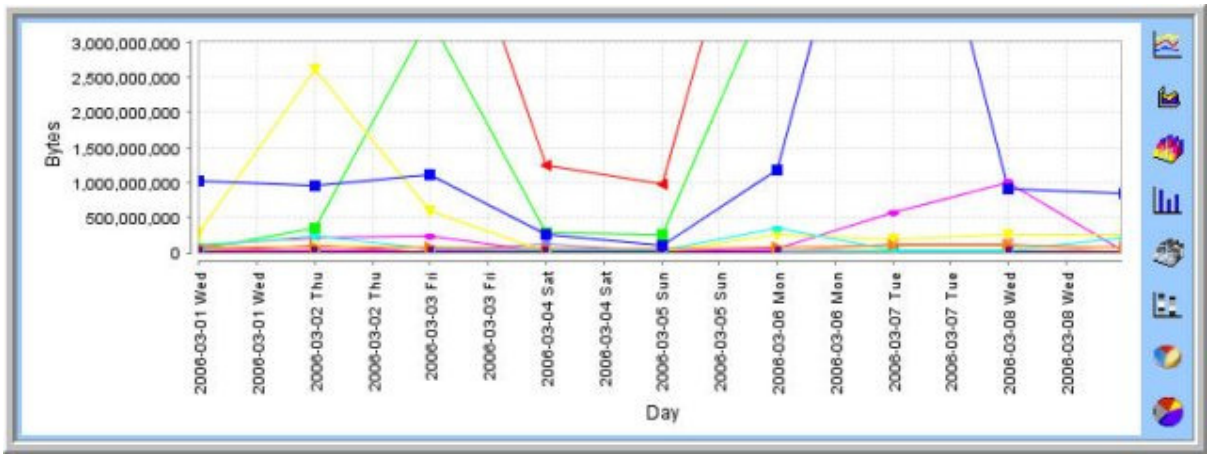
Step 1: Generate a report. Following is an example of a weekly baseline report.



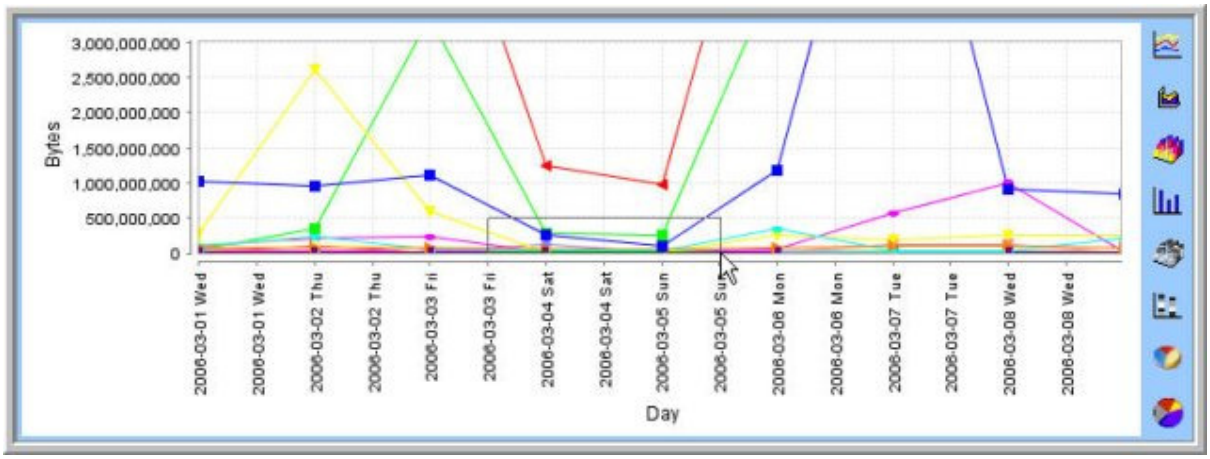
Step 2: Drag the moue to select an area within the graph into which you wish to drill down.

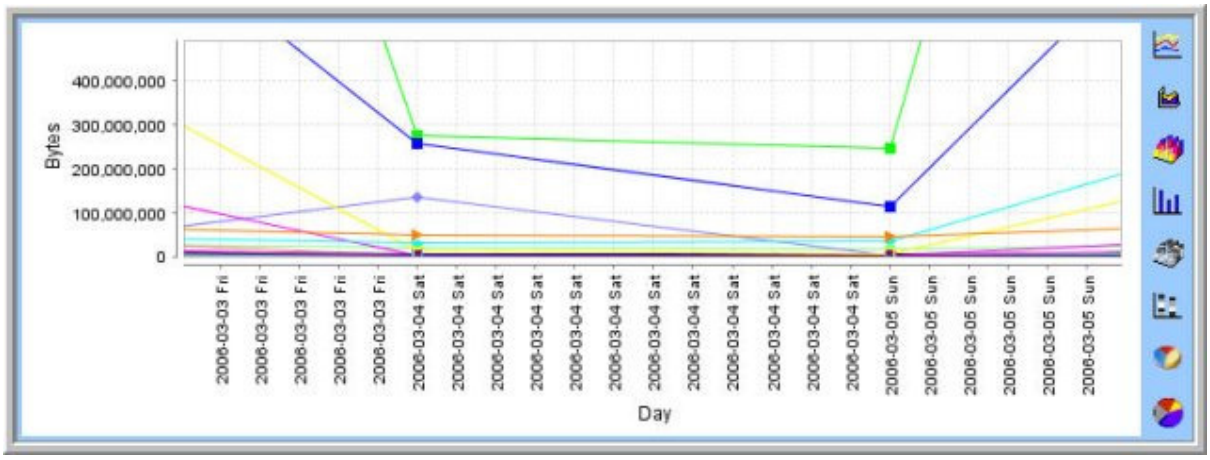


Step 3: A new graph is displayed elaborating only the selected area with greater detail.



Step 4: Repeat Step 2, as many times as you want.





2 - Drilling Down

The drill-down functionality of NetFlow Auditor gives the user single point analysis of traffic from the charting applet. All charts can be analyzed through the drill-down menu by using the left or right-click button of the mouse.

To drill-down into a chart:

Step 1: For line charts focus on a single point by clicking on a single point locater. For bar, area and pie charts skip to step 2. For charts with a lot of detail zoom in for more accurate drill-down analysis. See Zooming in.

Step 2: Right-click on the focus point/area to access the basic drill down menu or Left-click on the focus point/area to access the basic drill down menu.

Step 3: Select from the drill-down menu to analyze Devices/Business Groups/Applications/Time.

Step 4: Continue to drill-down on single points to further analyze traffic.

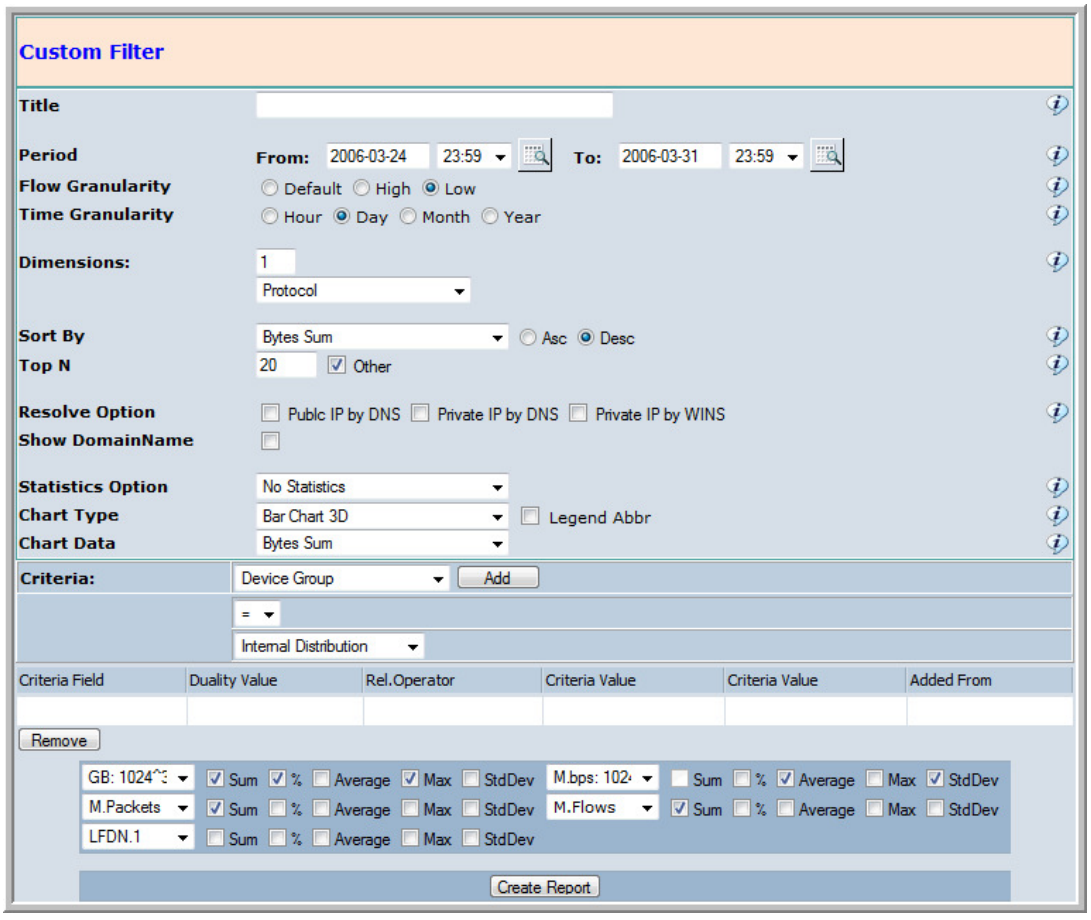
Part 3 – Traffic Filtering

Sometimes traffic is hidden within the network and is difficult to identify. The filtering capabilities of NetFlow Auditor enable the end user to further refine the traffic in the report in order to bring to the identify traffic patterns or trends for individuals, departments, or applications.

In order to begin first select one of the report starting points from the menu on the left of the NetFlow Auditor interface. For example, Long-Term/Usage/Applications/Protocol.

To begin filtering the report click the Filter button. 

The Filter Screen is now displayed in the Display Section and is populated with the previous report's data filtering criteria.



Change any of the properties as you see fit.

Title: Enter a title for your filtered report.

Period: Adjust the start date and/or End Date for the period of the data you wish to report on.

Flow Granularity: Leave as default or choose **High** for Real-Time (per minute or above) or **Low** for Long-Term (per hour or above).

Time Granularity: Choose a suitable time granularity based on the length of the report being produced.

> ***Note:*** For longer period reports ensure that you select an appropriately high level of granularity to reduce the time taken to generate the report. For example a report generated over a time frame of three months should be generated with a flow granularity of *Low* and a time granularity of *day*. Once the report is generated drill-down is still possible to discover greater detail.

Dimensions: Add objects you wish to display on your report. Two Dimensions results in a Stacked Bar chart. Time values create Comparative Baselines. Dimensions can be counted by using the Last Field Definition -LFDN- check box below. LFDN counts dimensions from the last dimension backward eg LFDN1, LFDN2 and so on. If LFDN is checked then the -Sort By- for LFDN becomes available.

Sort By: Sort has two options that become available as a result of earlier choices. -LFDN- sorts by count of respective dimensions (LFDN). -Field- is the Combined Dimension Name. Except for LFDN and FIELD, additional sorting can be performed after the report is produced. sorting performed after the report is created will only sort the -TopN- values requested.

Top N: Top 1 to Top N. Processing Capacity and time dependent. Use carefully when producing long comparative timelines for many multiples of results.

Resolve Option: -Public IP by DNS- reverse public IP to Domain Name by DNS server. -Private IP by DNS- reverse private IP to Domain Name by DNS server. -Private IP by WINS- reverse Private IP by WINS server. *Note this option may slow down the generation of the report depending on DNS server.*

Show Domain Name: Display domain name on report.

Statistics Option: -No Statistics- does not calculate a Time Chart series and is the fastest option. An option exists on the Report if a Time Chart is required. -Simple Statistics- produces the report with all the data. -Accurate Statistics- produces a full time series regardless of gaps and is useful for aligning service levels.

Chart Type: A variety of options exist for charting including; Time Chart, Stacked Area Chart, Stacked Bar Chart, Bar Chart and Pie Chart. Requires -Simple Statistics- or -Accurate Statistics- to produces a Time Series report.

Chart Data: As for -Sort-, -LFDN- is available after the check box is applied.

Criteria: Enter criteria from the menu and add as required. Multiple criteria are acceptable.

System Requirements

Operating System

- Windows XP (Standard Only)
- Windows Server 2003 (Standard Only)
- Centos Linux 4.4
- Red Hat Linux Enterprise 4
- Sun Solaris 8
- Sun Solaris 9

Database Engine

- MySQL 5.0

Web Server

- Apache Tomcat 5.5
- Java 2

Client

- Sun Java 2 Platform Standard Edition 1.5 or Higher
- Microsoft Internet Explorer 6 or Higher

Hardware

Table 1: System Requirements may vary depending on final configuration.

Users	Daily Traffic	CPU	RAM	HDD
1,000-5,000	50GB	2Ghz	1GB	80GB
5,000-10,000	100GB	2Ghz	1.5GB	80GB
10,000-20,000	200GB	Dual Core	2GB	80GB
20,000-50,000	500GB	Dual Core	4GB	120GB
50,000+	500GB+	Four Core	4GB	RFQ[4]

[4] Request for design quotation.