THE BIG BOOK OF NETWORK FLOWS FOR SECURITY

FlowTraq

TABLE OF CONTENTS

INTRODUCTION TO NETWORK FLOWS

Network flows represent the conversations that make your business work: emails, Web requests, VoIP calls, file transfers, and all the low-level back-and-forth that make a network a network. Among these conversations are also attacks: spam, scanning, malware, data exfiltrations, and other potential threats.

Network flows provide up-to-the-minute information about the communications taking place on the network, including who's sending how much data to whom, as well as how and when: IP addresses, port and protocol, exporting device, timestamps, plus VLAN, TCP flags, etc. This data is widely available from devices like routers, switches, firewalls, load balancers, hypervisors, and even as software to install on individual hosts. With data streaming in from multiple sources, a central location can get an excellent view of the network, including cross-border and purely internal traffic.

Cisco, the inventors of NetFlow, describe it as a phone bill for your network[1] — a listing of all the conversations that take place on your network, whether they last hours or milliseconds. Unlike an ordinary phone bill with hundreds of conversations, there are thousands or millions of "conversations" on the network at any one time. These conversations are the ebb and flow of data and control of a modern computer network, embodying the business processes that your network supports. The size of the conversation doesn't matter — a single-kilobyte communication can be as important to the operation of your network as a multi-gigabyte download.

[1] "Cisco IOS NetFlow and Security," Internet Technologies Division, February 2005

Cisco's NetFlow is just one of the many network flow protocols out there. Others include:

- J-Flow
- Cflow
- IPFIX
- sFlow®

Flow data provides granular insight into your network that can help you diagnose a potential problem. With the right tools, a network administrator can determine the source and destination of network traffic, prominent peering relationships, and common bottlenecks. Specifically, flow records contain information about:

- **Source IP address:** who is originating the traffic
- **Destination IP address:** who is receiving the traffic
- **Ports:** the application utilizing the traffic
- **Class of service:** the priority of the traffic
- **Device interface:** how the traffic is being used by the network device
- **Tallied packets and bytes:** the amount of traffic
- **And more:** including packet timestamps

THE DIFFERENT KINDS OF NETWORK FLOWS

NETFLOW SUITE OF PROTOCOLS

For the NetFlow suite of protocols — which includes IPFIX, Cflow, and J-Flow — we most often see version 5 (supported by the majority of devices), some combined v5/v7 (the Catalysts), and some version 9 on the newer devices. Don't be fooled by the ASA series of firewalls; they do not actually support version 9 flow exporting. Instead, these Cisco devices use NetFlow 9 to firewall events, similar to log lines: no real traffic records in there! NetFlow v5 uses a static packet format (and is in this way very similar to v7), defining IPv4 IPs, protocols, ports, and millisecond precision on flow start and end times. Version 9 uses a dynamic format, parsed based on a template which is sent around first. These templates are flexible and allow for expansion of the protocol in the future. Incidentally, IPFIX is also based on NetFlow and is versioned as NetFlow 10.

J-Flow and Cflow are the same as Cisco NetFlow v5. Only NetFlow v9 and IPFIX support IPv6.

NetFlow defines a "flow" as a unidirectional series of packets from IP A to IP B, using some protocol (TCP/UDP/ICMP/…). When the packets use either TCP or UDP, then the flow is further specified by a pair of ports; for instance, 10.20.30.40:53823->50.60.70.80:443 TCP. Often, since most communications require both sides to transmit packets, one will see NetFlow report two flows associated with every communication, accounting for the packets and bytes

that went in either direction. A proper flow collector and analyzer will correlate these with each other for you, so you can see a report of a full conversation. Most versions of NetFlow also support a sampling mode where only one in every N packets is used to update the flow counters. This is not very useful for forensic analysis of your network traffic, but helps keeps the CPU load on your router or switch down (see “To Sample or Not To Sample?” below). If full-fidelity NetFlow is required, consider using a SPAN, TAP or mirrored port, and generate NetFlow with a software tool or dedicated appliance without incurring the additional load on the router or switch.

The NetFlow suite of protocols is a powerful source of security and network debugging information. Since each and every network communication can be logged with millisecond precision, you can quickly determine who communicated with whom and when. Flow information is also much more tenable than raw packet data, allowing for a much quicker first look at the network. In other words, you can use flow data as a springboard to determine if further packet inspection is necessary.

SFLOW

The sFlow protocol is a completely different animal. Easily configurable through SNMP, its primary objective is to be a statistical network monitoring tool. Lots of different performance counters can be monitored through the sFlow protocol, and the biggest benefit of sFlow comes from its infinite scalability in large networks under heavy loads; however, this innovative statistical approach comes at a slight disadvantage in accuracy, granularity, and timing precision. To understand this, we have to take a closer look at how sFlow measures network traffic.

Unlike NetFlow, the sFlow protocol samples every Nth packet from the traffic stream, where N can be one-in-512, one-in-1024, etc. This means that some communications may slip by entirely undetected, and the sFlow collector software will not know about them. Larger communications, such as big downloads and online video content, will stand a much bigger chance of being reported, as there are many packets involved. The second drawback is the lack of accurate timestamping of the packet data. Sampled packets get forwarded as they are picked up from the datastream; however, they are not timestamped. Therefore, a small amount of uncertainty about the exact time of packet capture is introduced.

Although these tradeoffs render sFlow to be not particularly well suited to network forensic investigations (there is some statistical uncertainty as to when a communication began, how many packets were transmitted each way, their size, and when it ended), they are necessary to allow the virtually unlimited scalability that sFlow offers. If the network gets busy, it can fall back to a slower sampling rate, and keep load on the exporting device and sFlow collector down significantly.

WHICH FLOW SOLUTION IS RIGHT FOR YOU?

The answer to this question depends on the intended purpose of the implementation. At the ISP or large enterprise level, the hardware cost associated with tracking every communication through NetFlow is substantial, and can only be justified if the NetFlow data is used for security and network forensic analysis. If the goal is to simply get a rough overview of usage (“Who’s hogging my bandwidth?”), the sFlow protocol, or sampled NetFlow will suffice — it’s much more manageable and less costly. At smaller sites, the decision will usually be dictated by the switching and routing gear in the current network closet. Use what you have!

FLOW EXPORTERS

A flow exporter is a software or hardware engine that keeps track of all the current sessions that it "sees." In hardware, this might mean all the packets that a switch switches; in software, it's what can be seen on an interface or set of interfaces. It does this by maintaining a table of current sessions, called a conntrack (as in connection tracking) table. Typically, this table is scanned for updates at regular intervals, which are subsequently added to flow packets that are exported to a flow collector.

On very busy networks, sometimes an enormous number of sessions receive updates during the scan interval, thus resulting in a surge of flow packets to export. This sudden burst of output can overwhelm the UDP input buffers of the operating system on the collector, especially if the collector machine is heavily loaded or underpowered. This can lead to dropped flow packets and inaccurate, incomplete data. To combat this problem, you want your flow exporter to spread all the export packets over time, rather than send them in bursts. This smooths out the surges considerably, resulting in a steady stream of flow packets, even on very busy networks, and minimizes the chances of overloading a collector.

FLOW COLLECTORS

A collector is a server with software that can accept and interpret flow exports. Exporters send their flow summaries to collectors for storage and analysis. Most collectors summarize and aggregate the flows before storage, discarding the records. Although coarse, this approach is fastest, but the cost is the loss of forensic accuracy. Some collectors store all flow records, allowing full recall, and precise filtering. These full-fidelity analyzers are more powerful. For simply analyzing top network users and their top content, you may only need a flow aggregator. But because network flows are a summary, they can be stored compactly — a gigabyte of flow data can describe hundreds of thousands of gigabytes of actual traffic. This efficiency makes it suitable for full-fidelity recall, meaning every single record can be stored for later analysis, down to the smallest ICMP ping. With a full-fidelity network flow history, it becomes possible to perform intricate analysis of one's historical record, down to small transactions that occurred months ago. This enables an analyst to perform difficult tasks like tracking an intruder through multiple hops of SSH sessions, or separating potential DNS tunnels from ordinary requests. It also means that analysts don't have to be psychic — they don't need to know today what they'll need to search for tomorrow.

More than that level of intricate analysis, a fast full-fidelity system gives analysts an unparalleled feel for the network. Simply clicking around and exploring for a half-hour can give a far better idea of what to expect in the traffic than days in a lecture hall.

TO SAMPLE OR NOT TO SAMPLE?

On any given day, a typical networked host will send about 30MB and receive about 200MB. About 300,000 packets are switched. During peak times, the average workstation initiates two to four network UDP or TCP sessions per second, and each session averages 34KB in size, roughly 100 packets. What's more, these sessions are negative-exponentially distributed with regard to packet count. What does that mean? It means there are a lot more very short sessions of only a couple of packets than lengthy sessions with lots of packets.

When routers use sampling for network flow generation, an interesting thing happens. The sampling is done on a packet-count level, so a 1:512 sampling rate will grab roughly every 512th packet to update the flow state tables.

This is great for reducing CPU load. But it is not so great at reducing flow update rate. Here's why: With an average session size of roughly 100 packets, each sampled packet is very likely to be part of a flow that is not yet in the state table. This means an entry is created, which will lead to a flow update being sent. Compare this to 1:1 unsampled flow generation, where most of the packets will go toward updating existing entries in the flow state table. Flow state tables are typically exported when a flow is 60 seconds old, or the table is full, and the old ones need to be purged.

Leaving the exact math out for clarity, if unsampled flow generation results in a flow rate of X, then a 1:512 sampling results in a roughly 1/5th of the flow traffic being generated. Not 1/512th.

This is the intuitive answer, and the true results of sampling depend much on the precise mix of traffic present on the network. Also, some routers will use adaptive flow sampling rates to keep their flow export rates constant. This means that at busier times, the granularity of the data becomes less and less. Although this is nice for CPU time considerations on the router's end, it does not help much that the roughest data is collected during the heaviest attack! Sampling is simply not the correct approach to reduce cost or gain best visibility. You should design for 1:1 unsampled flow, because it builds in a safety margin during the biggest attacks.

NETWORK FLOWS FOR VISIBILITY

When collected in a central location and made available to view and search, network flows provide visibility into what's going over the network, including large-scale events like DDoS attacks and saturated links, and small-scale events like SSH logins and database connections. This allows an analyst to monitor for attacks, policy compliance, and data usage for billing. Whether it's a brute-force attack or on-the-sly Netflix watching, if it crosses the network, it's in the network flows.

A modern computer network encompasses a wide variety of Internet-capable devices, not all of them physical, joining and leaving your network in almost no time at all. The network infrastructure itself can change drastically with just a few keystrokes, without moving even a single wire of your physical infrastructure. A properly deployed network flow solution gives you exceptional visibility into your network, providing the best available balance between scope and depth. Network flows by themselves give you an excellent view, but combined with the proper analytical tools, it gives you unparalleled control over your network.

VIRTUALIZED NETWORKS

The ability to rapidly deploy virtual networks and virtual hosts has been a game-changer for many companies, allowing unprecedented flexibility. Network flows are fast enough to keep pace — the minute your virtual network goes live, so does your

flow export and you can see the flow and balance of traffic between VLANs at a glance.

MOBILE DEVICES

Mobile devices are particularly difficult to monitor because they move rapidly from network to network, are not easy to instrument, and do not have easily accessible log files. Network flows provide a convenient and robust means of monitoring these devices when they associate with your wireless networks.

REAL-TIME NETWORK MANAGEMENT

The more you know about your network, the better prepared you'll be for decisions you need to make:

- Is your Web service suffering a denial of service attack?
- Are all of your backups made on time, every time?
- That foreign IP address that is currently trying a brute-force attack against a system in your Chicago network — has it contacted any of your other networks today? Last week? Last year? Are they trying blindly or did they perform reconnaissance first?
- Can you reduce load on your Boston servers by moving functionality to San Francisco or do they experience peaks at the same time?

BANDWIDTH PLANNING

Whether your site is getting more popular, your team is expanding, your partners start preferring video conferences to voice calls, or the size of the average Web page is increasing, the only certainty is that you will need more bandwidth. But where? Long-term network

flow trends can tell you where resources are most likely to be needed most urgently. They can tell you which traffic is dominating your network and, at a glance, show you the rates at which it's increasing. This makes it easy to spot potential bottlenecks and strategically plan for expansion.

REGULATORY COMPLIANCE

Different organizations have different record-keeping requirements. For example, if you handle medical records, you are required to show HIPAA compliance; if you handle credit card details, you are required to track the flow of data to and from the system that processes this vital information, showing all access, even the smallest. If you monetize your network by hosting services, you need to rack bandwidth use accurately to determine usage billing. Relying on 95th-percentile billing can be risky; it can be gamed and you can miss spikes in traffic that degrade performance to other customers.

NETWORK FLOWS FOR SECURITY

Flow is a compact format, meaning it can be processed and analyzed much more quickly than a full packet capture format (fullcap). When analyzing network patterns for illicit behaviors, botnets, or data exfiltrations, the speed of analysis is key. If it takes hours to detect a potential data theft, and then a security analyst takes hours more to investigate, how many proverbial horses will have left the barn before the barn door is shut? Flow analysis delivers the power and speed to act quickly. At the same time, point-to-point communications are increasingly being encrypted, which dramatically and rapidly decreases the value of fullcap.

By analyzing flow data, network and security operations personnel can flag unfamiliar IP addresses (or IP addresses known to host malware), analyze distributed denial of service (DDoS) attacks, identify potential worms and botnets, track unusual data traffic patterns, find non-compliant users, and give you a detailed audit trail of all network activity. Flow analysis can also help you pinpoint unwanted data exfiltrations, identify causes of slowdowns, and spot where attacks or information leaks are coming from.

BOTNET PROTECTION

Once you learn your organization's typical network patterns, you can use your flow data to help detect network anomalies that often indicate botnets, rogue servers, unauthorized clients, or other network threats.

DDOS AND BRUTE FORCE ATTACK DEFENSE

A significant traffic surge can indicate a potential DDoS or brute force attack, two of the most common network threats around. DDoS attacks can make your websites, email, or other applications unavailable to your customers or employees, while a successful brute force attack allows a perpetrator, virus, or worm to penetrate your network with compromised credentials. By regularly analyzing your flow traffic you can recognize these types of malicious threats in real time so you can immediately react to them and select the appropriate defense.

WORMS, SCANS, AND NETWORK RECONNAISSANCE PREVENTION

Worms propagate through your network by rapidly looking for hosts with common vulnerabilities and exploiting those weaknesses to spread throughout your network. By using flow data, this scanning behavior can be easily identified when several internal systems show the same bad pattern in rapid succession. This is a very common reconnaissance technique used by attackers, and can serve as an early warning for other more malicious attacks to come. Flow data can help you catch and control these incidents before they become a real problem.

DATA EXFILTRATION RECOGNITION

You may already spend time and resources searching for viruses on computers, blocking spam, and tracking down abuse. But it's not only worms and viruses that may be exfiltrating your most sensitive data, it could be anyone within your own walls. By monitoring flow data to detect anomalies, you can

immediately recognize undesired uploads and data breaches from your network. This allows you to stop a leak before you face millions of dollars of cyber theft damages, lawsuits and public embarrassment, or before your CEO, CSO or CIO lose their jobs due to failed data security practices.

DATA BREACH DEFENSE

To spot potential data leaks or information security breaches, you can analyze network flow records to quickly recognize hosts initiating a connection, receiving data outside of normal thresholds, or exhibiting unexpected network behavior patterns. With the proper tools, flow data analysis can help you keep sensitive information safe from outside intruders and unauthorized insiders.

WHICH NETWORK FLOW IS BEST FOR DDOS DETECTION AND MITIGATION?

Successful DDoS triage and mitigation depend on two things: *speed* of detection and *accuracy* of detection. Users considering a DDoS solution often ask if it is best to use NetFlow or sFlow. To understand what is best, it's important to understand the differences between the two types of flow data.

NetFlow (and the very closely related cFlow, JFlow, and IPFIX) is a summary record format, where a router or other exporting device tabulates the statistics on each flow of packets flying by. A flow is typically defined as the 5-tuple of sending IP and port, receiving IP and port, and the protocol. Each packet is tabulated and added to the appropriate row in the table: 1 more packet, X more bytes. State is kept on every flow that the exporter observes, and when a flow is 60 seconds old, the record is sent to the collector, which has the task of detecting the denial of service attack.

sFlow takes a slightly different approach, keeping no state at all. Instead sFlow randomly grabs one in every N packets flying by and immediately sends it to the collector. Although this approach may appear somewhat less accurate than the NetFlow tabulation, it is actually very good for fast DDoS detection. As the flood or amplification attack starts to ramp up, the rate of packets flowing by the exporter starts to increase very rapidly. This means that the number of packet samples going to the collector (which is responsible for the DDoS detection) starts to increase immediately.

Although the NetFlow approach ensures that no packet is missed, which is great for accurate network forensics, the nature of the export timer may result in much slower detections. If the NetFlow exporter has sufficient memory to keep state on all the attack traffic, it may take up to 60 seconds before the detecting collector sees any evidence of the attack! Thankfully though, many newer NetFlow-capable devices can be tuned to export at higher rates, resulting in improved detection times.

Detection accuracy is another matter. Since both sFlow and NetFlow transmit information on sending and receiving port and both transmit information on flag combinations and IP addresses, it is really up to the collector to make an accurate detection. Distinguishing a DNS or NTP amplification attack from a SMURF, or a FRAGGLE attack from a SynFlood, is key in performing effective mitigation. Most triage scenarios (whether using a scrubbing device or manually mitigating) rely on knowing a couple of key factors:

1. What are the targets being hit?
2. Are the bit/packet rates sufficiently high to impact the service?
3. What is the specific type (or types) of attack?

Both NetFlow as well as sFlow provide sufficient detail to accurately make that determination if the detection logic is present in the collector software.

In practice, both sFlow and NetFlow variants can be deployed very successfully in DDoS detection. Although some sFlow deployments detect DDoS attacks in as little as 3 seconds, the nature of the Internet has made it such that it can take some time before the attack reaches full strength. Combined with the

faster NetFlow exporters that have recently started to reach the market, the speed advantage of sFlow is starting to fade. Also, keep in mind that many environments will not have a choice, as the exporting hardware is already in place, supporting only a single type of export. So the choice may not be yours, and both approaches can be used to tune and finesse for detection speed and accuracy. Most importantly, if you have both, then use both. The better the visibility, the better your defenses.

THE SCALABILITY OF NETWORK FLOWS

Flow analysis will get even faster in the future. Although packet volumes are growing exponentially, network flow volumes are not growing as quickly. How is this possible? Individual flows are getting bigger (meaning more packets per flow) and this is because network sessions are primarily driven by user behavior, such as retrieving email, browsing Web pages, and consuming media. For example, we can only watch one movie at a time and read one email at a time. Analysis speeds will therefore grow over the years, as the time to process a single flow remains constant, regardless of the packet size of that flow.

Moving flow analysis into the cloud delivers additional advantages. We are able to harness the processing power to deal with transient, high-volume threats. Flow volumes may increase one-hundred-fold during a DDoS attack, which can put a strain on right-sized in-house flow processing equipment. In the cloud, flows are analyzed on a much larger platform than they would be in house; CPU resources are available to quickly detect, analyze, and mitigate these behaviors without resource-constrained slowdowns. Some crafty attackers and very stealthy bad behaviors only become apparent when viewed over many customer networks at once. The final benefit of analysis in the cloud is the most exciting of all — by analyzing flow data from thousands of sources all over the Internet together in a single window, emergent behaviors become visible that would otherwise be hidden from view when analyzed in isolation in a single network. This powerful benefit is unlocked by moving flow data to the cloud.

TIPS ON CONFIGURING FLOW EXPORT

For tips on configuring flow export from a variety of routers and switches, check out the following resources:

- Flexible NetFlow export from Cisco routers
- Simple NetFlow export from recent Cisco routers
- NetFlow export from older Cisco routers
- J-Flow export from Juniper SRX Series routers
- NetFlow export on VMWare vCenter with ESXi
- NetFlow export on Open vSwitch SDN

ABOUT FLOWTRAQ

FlowTraq® provides software and services for high-performance network monitoring, analytics, security and forensics to detect a range of network behaviors, including distributed denial of service (DDoS), brute force attacks, botnets, worms, network scans and other network traffic anomalies. FlowTraq is compatible with all the common flow formats — including NetFlow, sFlow®, IPFIX, Cflow, J-Flow, and PCAP — whether sampled or not. And thanks to a full-fidelity parallel database, FlowTraq can recall every individual traffic flow that crossed your network, no matter how long ago. This unlimited filtering capability is necessary to find an individual communication that is the source of the malware, control of the botnet, or destination of exfiltrated data.

To learn more about FlowTraq, visit **www.flowtraq.com**. Experience FlowTraq for yourself with a free 14-day trial (available to qualified organizations). To request a trial, visit **www.flowtraq.com/trial**.

FLOWTRAQ

16 Cavendish Court, Lebanon, New Hampshire 03766
Phone +1 (603) 727-4477 | Email info@flowtraq.com