SUNIL AMIN, CISCO STEALTHWATCH

FIRST TC 2017, AMSTERDAM

SECURITY ANALYTICS WITH NETWORK FLOWS

AGENDA

- What is a Network Flow?
- Why are Network Flows valuable?
- Preparing Network Flows for Analysis
- Analysis Use Cases
- Tools
- Q&A

WHAT IS A NETWORK FLOW?

MY DEFINITION

A record of a unidirectional IP(L3) network communication between two L3 endpoints during some time period.

Contains, at a minimum, the 5-tuple extracted from the IP packet header and associated timestamps.

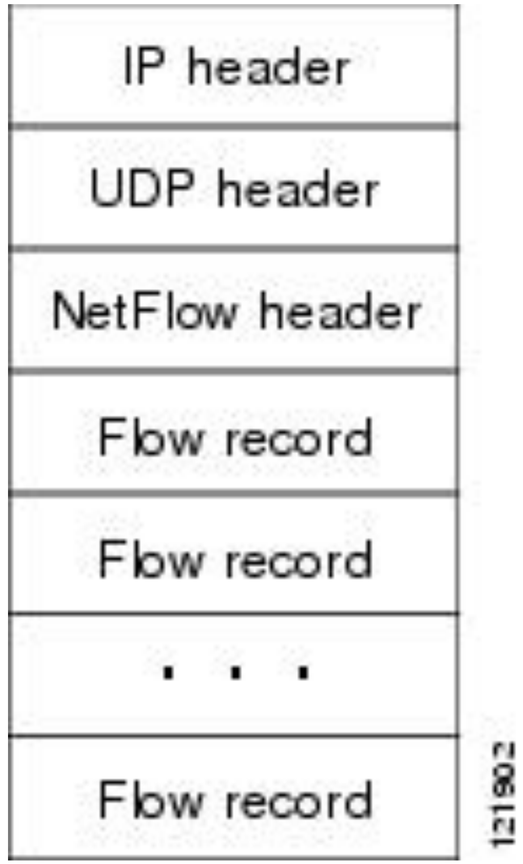
Start Time	Finish Time	Source IP Address	Source Port	Destination IP Address	Destination IP Address	IP Protocol
2010-09-01 00:00:00.459	2010-09-01 00:00:04.345	10.202.1.1	24920	10.202.100.100	80	TCP
2010-09-01 00:00:00.459	2010-09-01 00:00:04.345	10.202.100.100	80	10.202.1.1	24920	TCP

POSSIBLE DECORATION

Category	Fields
Traffic Volumes	L3 Byte Count, L3 Packet Count
TCP	TCP Flags
Network Device	Router or Switch Interface
L3 Routing	Next Hop, Source and Destination prefix mask, Source and Destination Autonomous System Numbers
Firewall	Firewall Rule and Action, User
L7 Application	HTTP Headers, DNS Request/Response
Many more ...	

NETFLOW & IPFIX

- Cisco NetFlow
  - Introduced as a traffic accounting and troubleshooting tool for switches and routers
  - v5 and v9 the most common
  - RFC 3954 (“Informational”)
- IPFIX (~NetFlow v10)
  - IETF Standards Track
  - RFC 7011
  - Broad network infrastructure vendor support

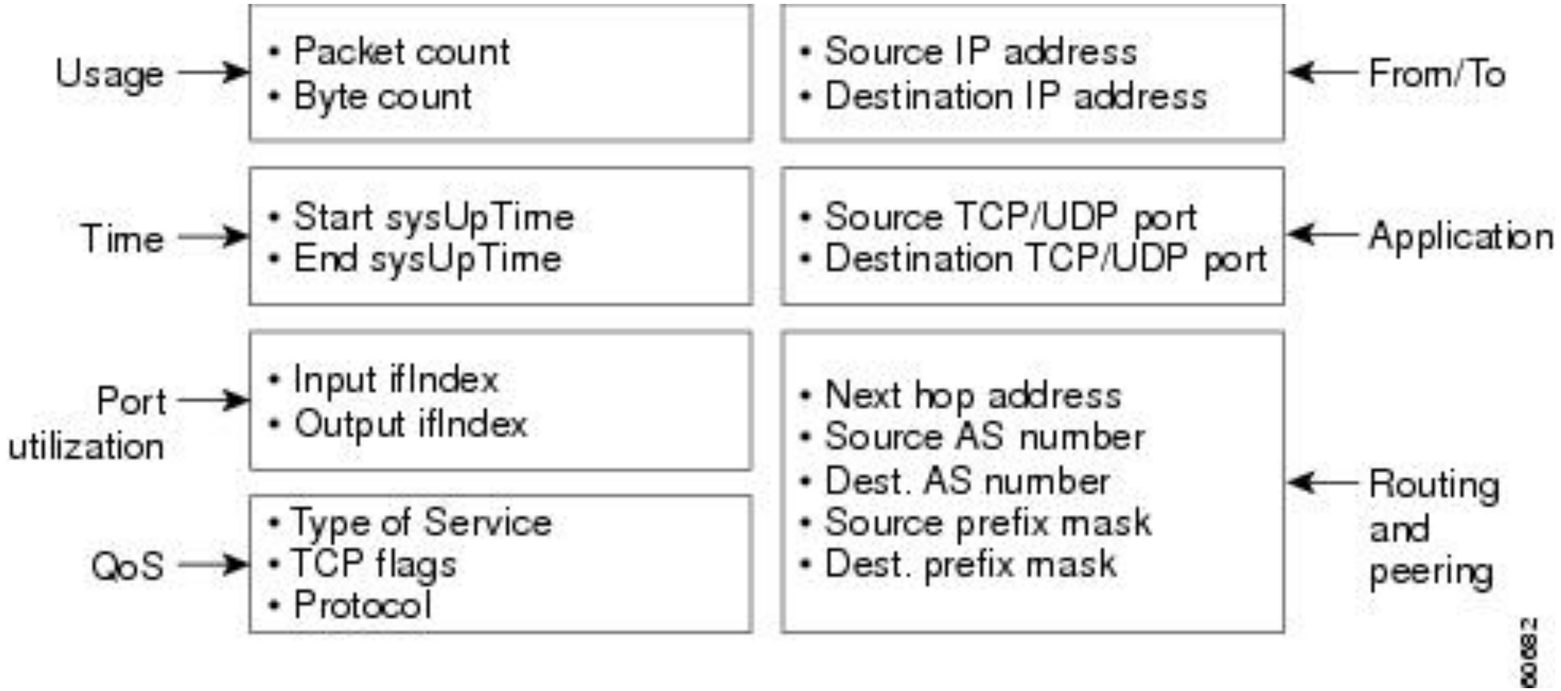


MOST COMMON SOURCES

- Inline Network Infrastructure
  - Most devices passing packets in your network
- Passive Software “Generators”
  - Other sources of packets
  - Endpoints
  - Network Tap or SPAN port

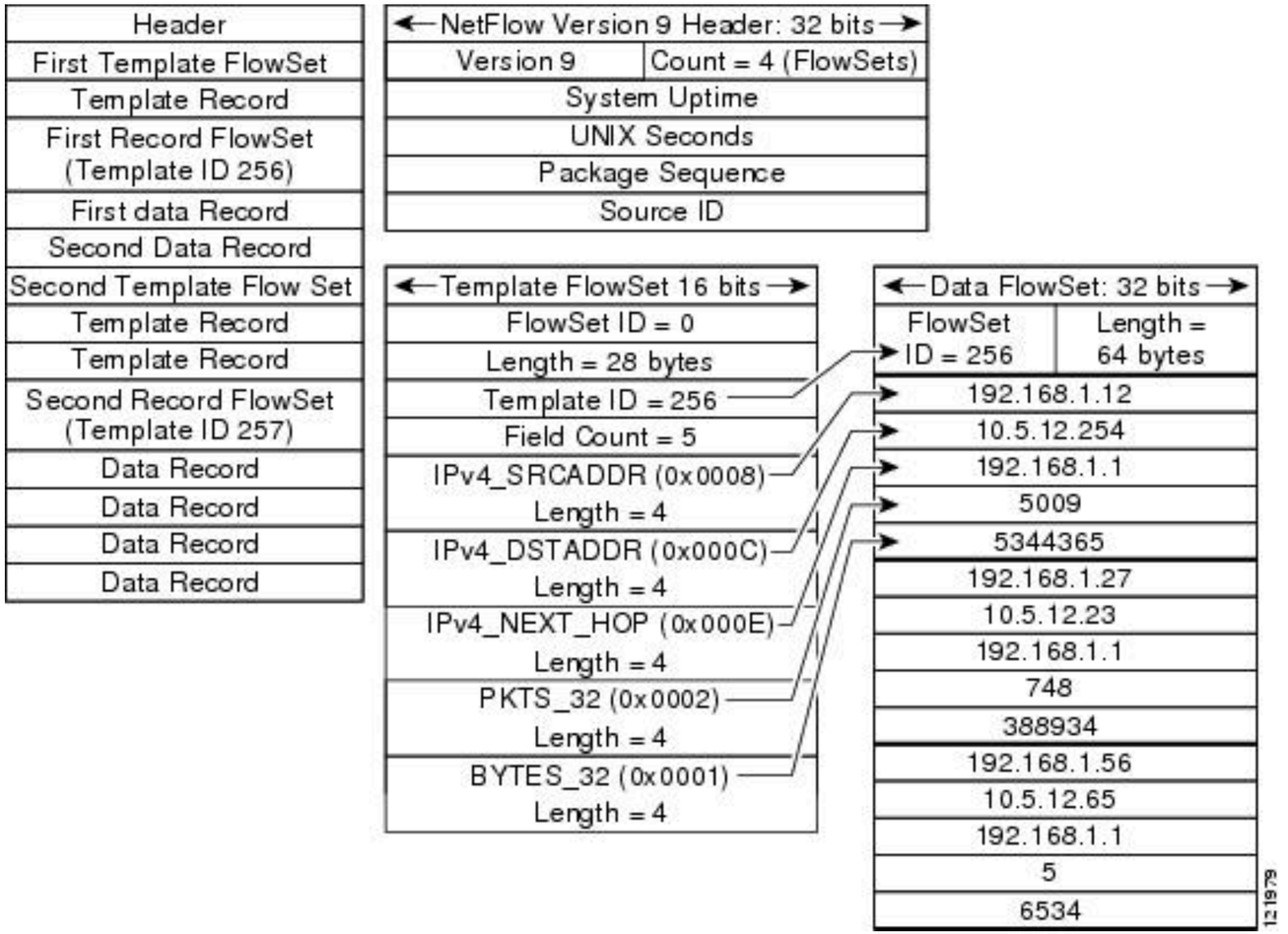
NETFLOW V5

- Fixed Content
- IPv4 Only



NETFLOW V9

- Dynamic Content
  - Runtime “Templates”
- 100+ Cisco defined fields
- Allows for vendor extensions



IPFIX

- Subtle structural differences with NetFlow v9
- Dynamic Content
  - Runtime “Templates”
- Allows for variable-length fields e.g URLs
- 450+ IANA defined fields
- Allows for vendor extensions

OTHER VARIATIONS – MORE NETWORK INFRASTRUCTURE

- JFlow - Juniper Networks
- Cflowd - Juniper/Alcatel-Lucent
- NetStream - 3ComHP, Huawei
- RFlow - Ericsson
- AppFlow - Citrix
- sFlow - Many vendors

SAMPLED NETWORK FLOWS

Beware: Not Complete!

(1) Deterministic: One packet in every n packets or
(2) Random: One packet randomly selected in an interval on n packets

SYNTAX VS. SEMANTICS

Warning!

Not everything carried in NetFlow v9 or IPFIX is a Network Flow.

Beware of events!

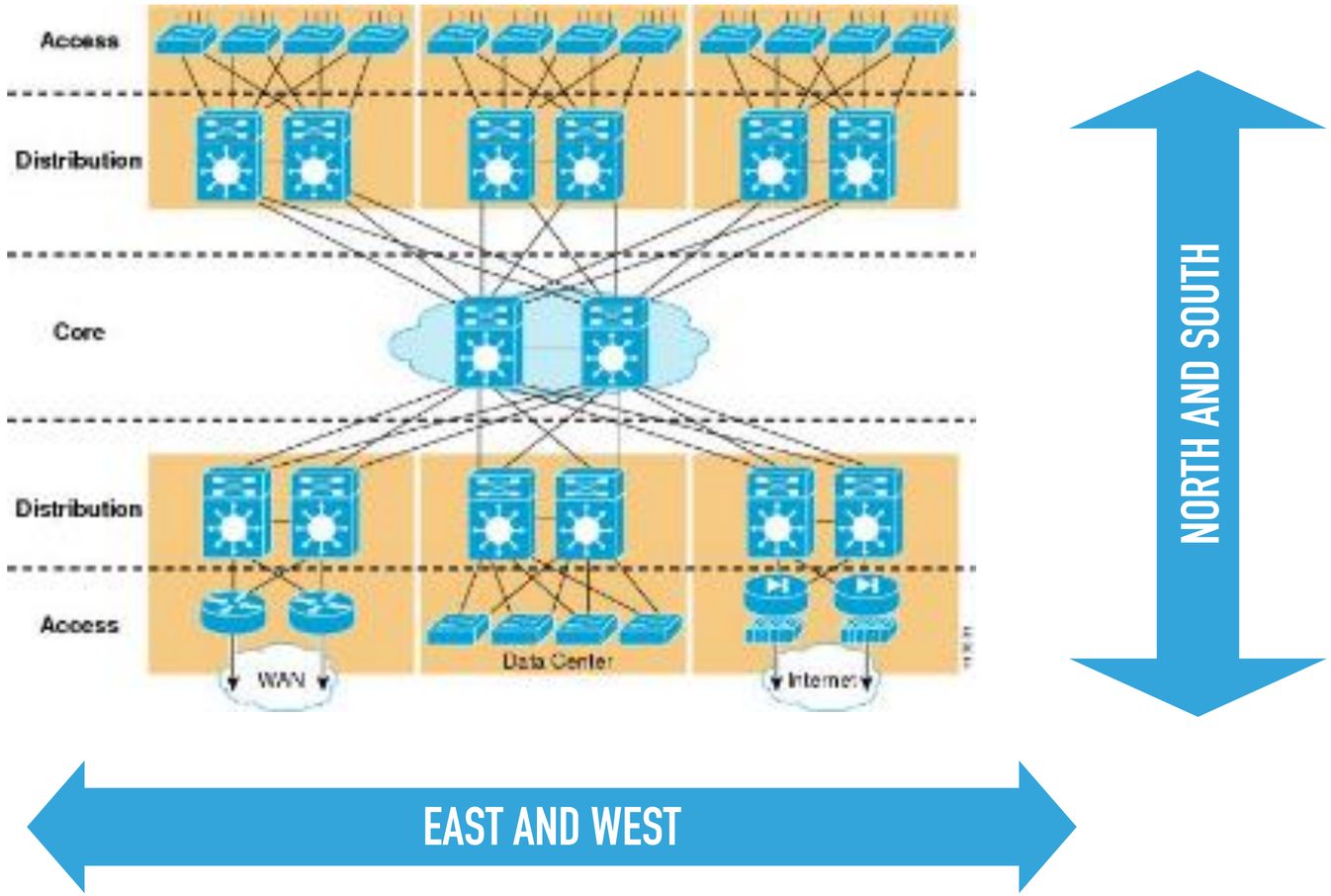
NOT JUST NETFLOW OR IPFIX

Field	Description
version	The VPC flow logs version.
account-id	The AWS account ID for the flow log.
interface-id	The ID of the network interface for which the log stream applies.
srcaddr	The source IPv4 or IPv6 address. The IPv4 address of the network interface is always its private IPv4 address.
dstaddr	The destination IPv4 or IPv6 address. The IPv4 address of the network interface is always its private IPv4 address.
srcport	The source port of the traffic.
dstport	The destination port of the traffic.
protocol	The IANA protocol number of the traffic. For more information, go to Assigned Internet Protocol Numbers.
packets	The number of packets transferred during the capture window.
bytes	The number of bytes transferred during the capture window.
start	The time, in Unix seconds, of the start of the capture window.
end	The time, in Unix seconds, of the end of the capture window.
action	The action associated with the traffic: • ACCEPT: The recorded traffic was permitted by the security groups or network ACLs. • REJECT: The recorded traffic was not permitted by the security groups or network ACLs.
log-status	The logging status of the flow log: • OK: Data is logging normally to CloudWatch Logs. • NODATA: There was no network traffic to or from the network interface during the capture window. • SKIPDATA: Some flow log records were skipped during the capture window. This may be because of an internal capacity constraint, or an internal error.

WHY ARE NETWORK FLOWS VALUABLE?

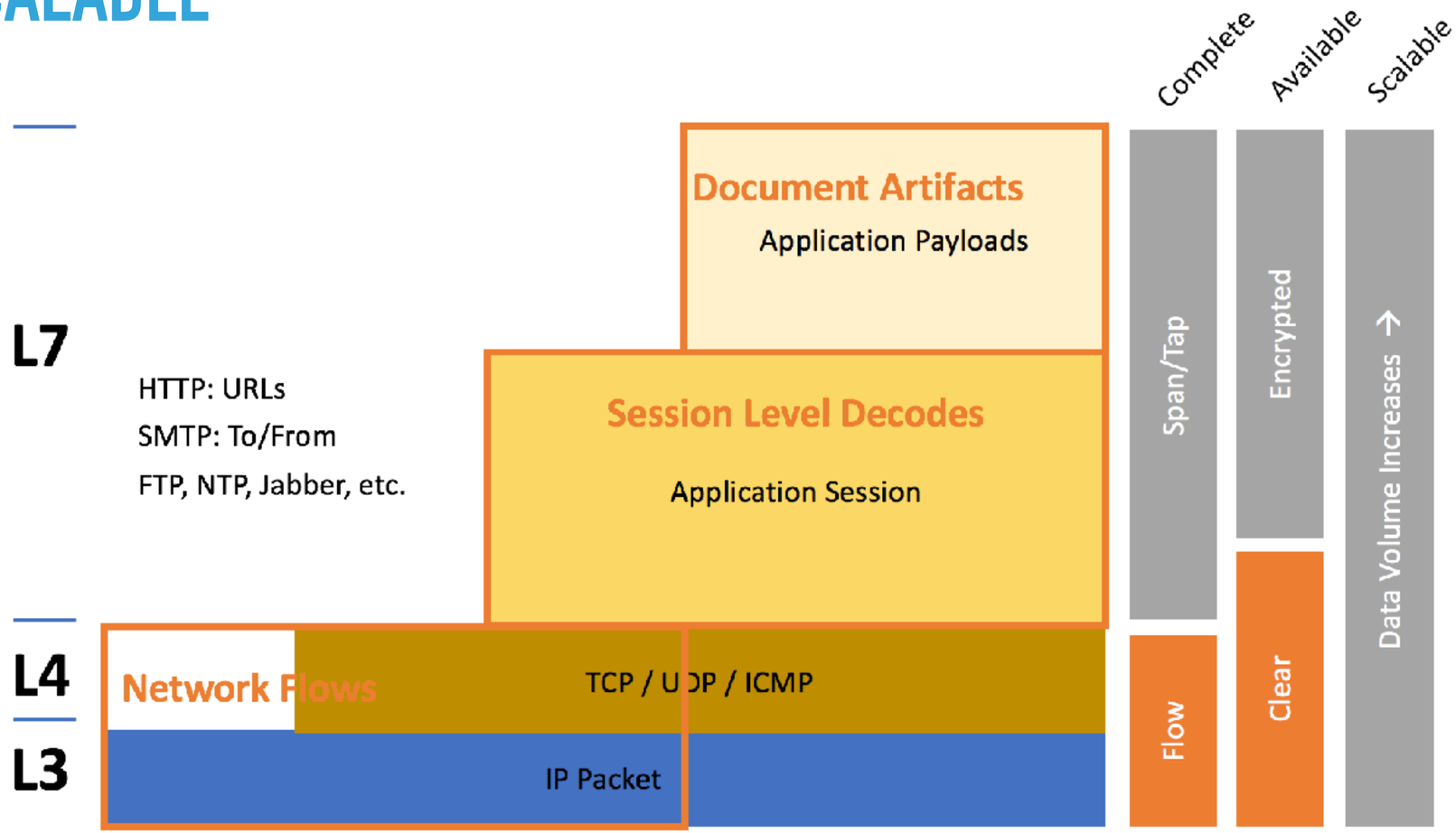
THEY PROVIDE THE "GENERAL LEDGER"

- Visibility beyond the perimeter
  - Not just North-South
- Internal Threats
  - East-West



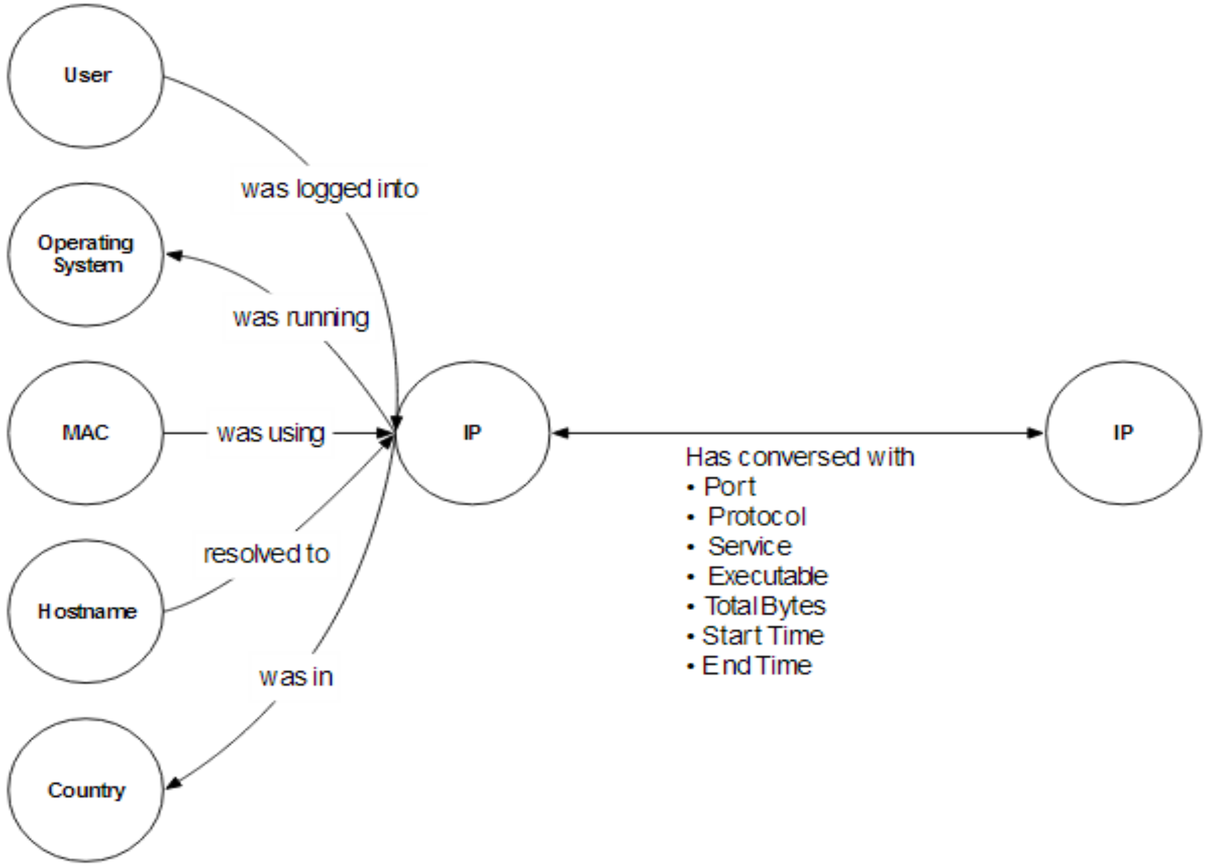
COMPLETE, AVAILABLE, SCALABLE

- Complete
  - Sources pervasive within the network
- Available
  - Headers always in the clear
- Scalable
  - 1-5% of traffic volume



CONTEXT AND CENTRICITY

User centric
Application centric
Host centric
Geolocation centric
IoC centric
Incident centric
Threat actor centric
File centric
File change centric
Vulnerability centric
Business process centric
Tag centric
Domain name centric
Session error centric

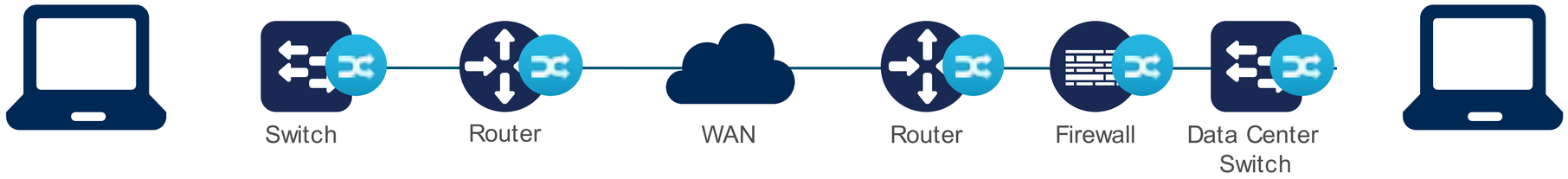


PREPARING NETWORK FLOWS FOR ANALYSIS

MAKE ANALYSIS EASIER

(1) De-duplication
(2) Bi-flows
(3) Correlation over time

DE-DUPLICATION



Every device a conversation traverses will report the same unidirectional Network Flow.
Inevitable as coverage extends to all possible routes through the network.

(1) Compress: keep one copy of the common fields. e.g. IP Addresses, ports
(2) Do not discard data: merge other fields into one record
(3) Avoid misreporting volume: select one device for counts (manual, first-reporter, max)

DE-DUPLICATION

source	srcaddr	srcport	destaddr	destport	proto	app
s1	10.202.1.1	24920	10.202.100.100	80	TCP	http
s2	10.202.1.1	24920	10.202.100.100	80	TCP	
s3	10.202.1.1	24920	10.202.100.100	80	TCP	

sources	srcaddr	srcport	destaddr	destport	proto	app
s1, s2, s3	10.202.1.1	24920	10.202.100.100	80	TCP	http

BI-FLOWS

Correlate unidirectional flows into bidirectional flows using addresses and ports where possible.

Determine initiator (client) using manual or heuristic techniques.
e.g. know server ports, lower port is server, first seen, TCP flags
This can be hard but very valuable!

If you are lucky: the source implements RFC 5103

BI-FLOWS

srcaddr	srcport	destaddr	destport	proto	packets	octets
10.202.1.1	24920	10.202.100.100	80	TCP	5	1025
10.202.100.100	80	10.202.1.1	24920	TCP	17	28712

clientaddr	clientport	serveraddr	serverport	proto	clientpackets	serverpackets	clientoctets	serveroctets
10.202.1.1	24920	10.202.100.100	80	TCP	5	17	1025	28712

CORRELATION OVER TIME

Most source will splice long running flows in to segments (“Active Timeout”)
Combine these segments to form “Complete” flows.
Keep a copy of the segments so as to not loose the temporal information.

Flow end can be determined by:
(1) For TCP: FIN flag seen, inactivity (“TCP Inactive Timeout”)
(2) For UDP: inactivity (“UDP Inactive Timeout”)

ANALYSIS USE CASES

KNOW WHAT IS ON YOUR NETWORK – GENERATE WHITELISTS

- Discover internal address space:
  - Look for flows where both endpoints are not on the internal whitelist (start with RFC 1918)
  - Update the whitelist
- Discover internal services:
  - Look for flows where the server is internal and group by the port/protocol. Exclude servers on service whitelists.
  - (1) Update the whitelist or (2) you have a rouge server

MORE WHITELISTS

- Keep your firewall honest:
  - Look for flows where the client is internal and the server is external and the port/protocol is blacklisted by your firewall
  - e.g. External SMB servers
- Look for blacklisted services
  - Look for flows where the server is serving Telnet or other out-of-policy service

REVEAL RECONNAISSANCE

- Address Scans
  - Look for flows from a single client to more than X servers within a Class C address range within Y seconds
- Port Scans
  - Look for flows from a ingle client to more that X ports on a single server within Y seconds
- Exploitation
  - If any of the above flows get a response from the server (serverpackets > 0)

BAD BEHAVIORS

- Reverse Shell
  - Look for flows where the server is on 22/TCP and the server/client byte ratio is high
- Brute Force Login Attempt
  - Look for flows where the client attempts multiple connections to the same Server, over the same port/protocol with small packet counts

ANOMALY DETECTION – STATISTICAL MODELING

- Data Hoarding
  - For internal hosts, generate a time-series of bytes received
- Data Exfiltration
  - For internal hosts, generate a time-series of byte sent, as a client, to external hosts

PUTTING IT ALL TOGETHER

- Initial IOC (X): Waterhole campaign targeting the client's industry has been disclosed
- Search the General Ledger: Reveals an internal host (A) that accessed the disclosed site as cross-referenced with passive DNS.
- Narrow the search: Retrieving all the flows immediately following the time of the access
  - Found HTTP connections to external host that had never been seen before (Y) - good candidate for drive-by download
  - .. followed by SSH reverse shell to external server (Z)
  - .. followed by address scanning on port 445 (SMB) and 135 (MS-RPC)
- SSH Server Z is now a new IOC .. rinse and repeat

TOOLS

THE BASICS

- SiLK from CERT NetSA
  - Collects, stores and process NetFlow v5, v9, IPFIX
  - Many unix tools including PySiLK
- nfdump
  - Collects and stores NetFlow v5, v9
  - Limited processing
- ntop
  - High performance NetFlow and IPFIX capture and generation tools.
  - Free for non-commercial use. Some tools are commercial only.

BIG DATA

- ELK Stack
  - Logstash has an NetFlow/IPFIX input plugin
  - Elasticsearch for search indexing
- Apache Spot
  - Full cybersecurity “big data” stack
  - Hadoop, Kafka, Spark
  - NetFlow v5/v9 support via nfdump
  - DNS request/response from packet captures
  - Machine Learning platform

QUESTIONS?