Zurich Research Laboratory

Dynamic Adaptation of Flow Information Granularity for Incident Analysis

Marc Ph. Stoecklin <mtc@zurich.ibm.com>
Andreas Kind <ank@zurich.ibm.com>
Jean-Yves Le Boudec <jean-yves.leboudec@epfl.ch>

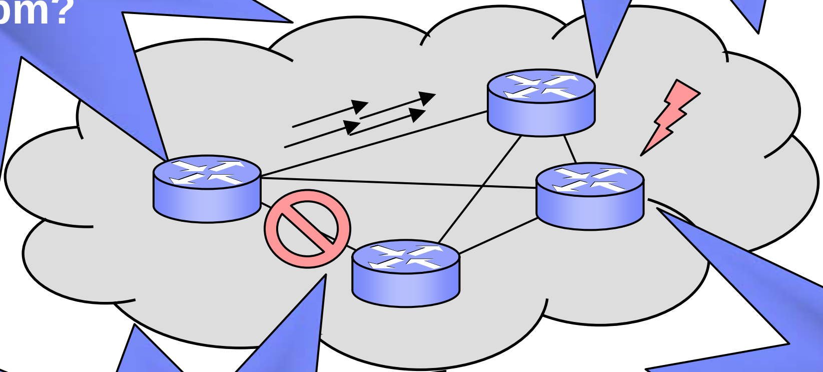
Jan 9, 2008 | FloCon2008

Outline

- Problem statement and objectives
- Adapting flow information granularity
  - Increasing granularity with Zoom Monitors
  - Decreasing granularity with lossy compression
- Implementation
- Results
- Conclusion and outlook

What caused the surge of traffic to the mail server last Tuesday at 2pm?

What is currently going on in my network?



What is causing this abnormal network activity?

What was the sequence of events before the incident?

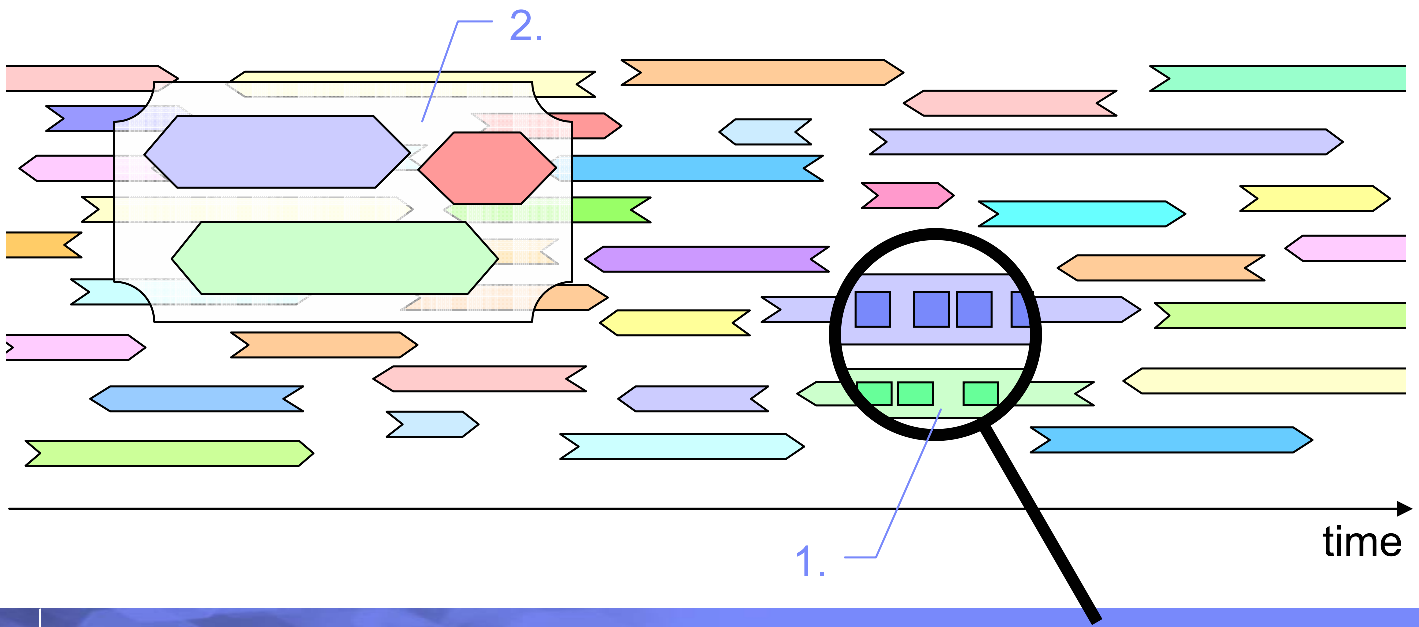
Problem Statement

- Trade-off in network **traffic information collection** for **incident analysis**
  - **Raw packet traces**: finest level of detail but impractical to manage and search
  - **Flow traces**: high-level traffic abstraction but aggregated
- Traditional flow exports may **not provide traffic details required** to understand causes of incidents
  - Missing layer 3 and layer 4 header information
  - No packet content information
- Flow-level information is still a **considerable amount of data**
  - Flow record collections are still tedious to search, store, and analyze
  - Majority of this (raw) information is never accessed

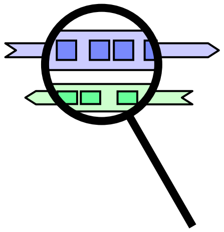
Objectives and Goals

- Extend a collector system to provide more accurate incident analysis
- Adapt information granularity depending on relevance of the traffic:

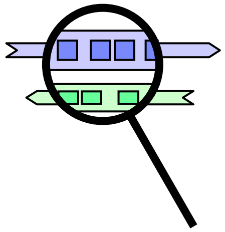
1. Focus in on particular traffic events to obtain more details
2. Compress known/less relevant traffic events (conserve a meaningful abstraction)



Increasing Traffic Information Granularity

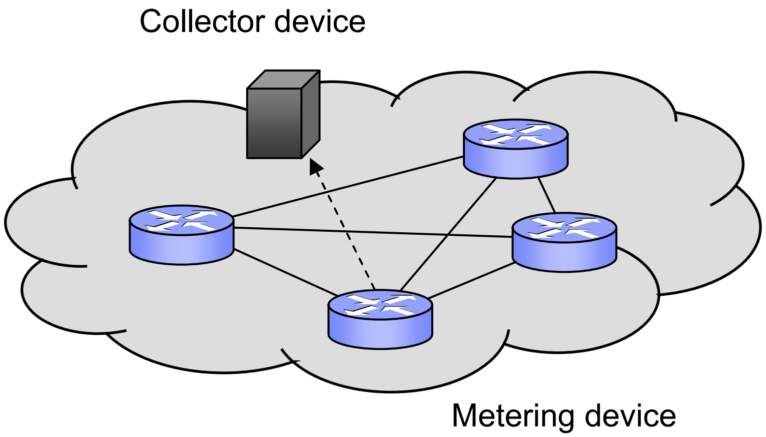


- **Problem**
  - Collecting detailed traffic information is cumbersome
  - Fixed and limited amount of information in default flow exports (e.g., NetFlow v5)
    - Valuable information may have been lost along with flow aggregation
- **Traditional approach (on-going anomaly)**
  - Physically attach a probe or packet dumping device at router (e.g., tcpdump with filtering)
  - Collection of rigid traffic information (e.g., entire packets): complex analysis
- **How to simplify data collection? Create Zoom Monitors!**
  - Dynamically controlled collection of traffic information at desired level of detail
  - Central management console for coordination
  - Make use of capabilities of network device inventory (routers, switches): reporting/dumping



Zoom Monitors

- **Specification**
  - Metering point and collector device
  - Zoom monitor lifespan
  - Filter criteria
  - Traffic aspects to be exported



- **Export collection and display**
  - Reconfigure metering device to create specific exports
  - Prepare collector device to store exported traffic information
  - Centralized management and display

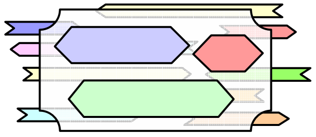
- **Examples**
  - Show me the payload of all DNS requests of host 10.3.4.5 during the next 10 minutes
  - Look for all internal hosts scanning on TCP service port 9996 (e.g., candidate worm traffic)
  - Inspect GET/POST requests and virtual servers accessed on web server 10.4.5.6
  - Export unsampled flow measurements from subnet 10.9.3.1/24

Decreasing Traffic Information Granularity

- **Problem**
  - Most stored traffic information is irrelevant for incident analysis (never accessed)
  - Redundancy (limited value): Increased storage overhead and search complexity

- **Traditional approaches**
  - Rolling database (FIFO): keep all records up to a limit (e.g., #records, age): information removal
  - Uniform summarization: adapt resolution of information (hourly, daily, weekly)
  - Keep top-k entries (according to some aspect)

- **How can we do better?**
  - Majority of network events is known or recurring
  - Gradually compress information of irrelevant traffic events in a lossy fashion
    - With minimal impact on incident analysis tasks
  - Summarize similar events (coarse-grained representation)

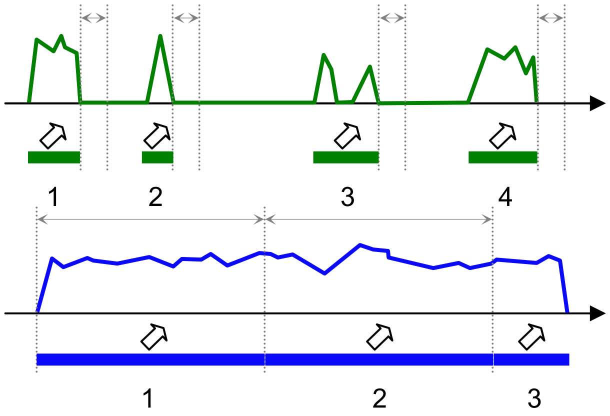


⬀ = exported flow record

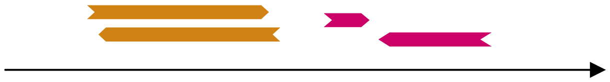
↔ = inactive/active timeouts

Observations

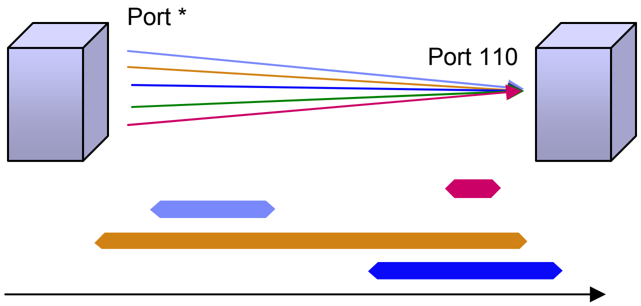
- **Flow exports**
  - Multiple exports for a single connection
  - Examples:
    - Long-lived connections (streams, remote sessions, etc.)
    - Timeouts on routers (inactive/active timeout)



- **Bi-directionality**
  - Most flows have a reversed counterpart



- **Information similarity**
  - Sets of records with limited added value on the flow level
  - Groups of flows with similar properties (Web, mail, printer traffic, polling)
  - Uniqueness: ephemeral port, time stamps, byte and packet counters



Compression Model[1]

	Abstraction models			
	Flow record	Flow	Conversation	Session
Raw exports	Yes	No	No	No
Flow definition	Yes	Yes	Yes	No (subset thereof)
Direction	Uni-directional	Uni-directional	Bi-directional	Bi-directional
# Flow records	1	≥ 1	≥ 1	≥ 1
# Flows	1	1	1 or 2	≥ 1 or ≥ 2
# Conversations	1	1	1	≥ 1

[1] without prior knowledge such as domain or application specific information

Implementation

- **Metering device configuration for Zoom Monitors**
  - Reconfiguration of metering devices
  - Management console
- **Export collector**
  - Collection and storage
  - Traffic information compression
  - Data querying

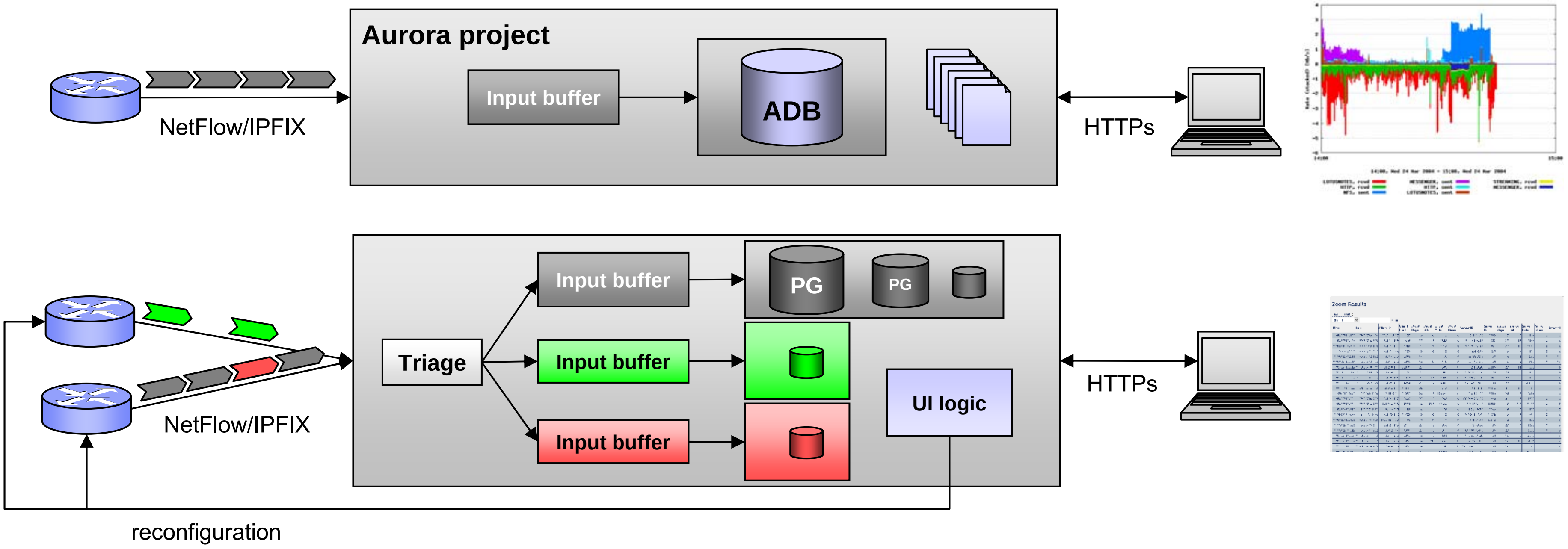
Metering Device Configuration

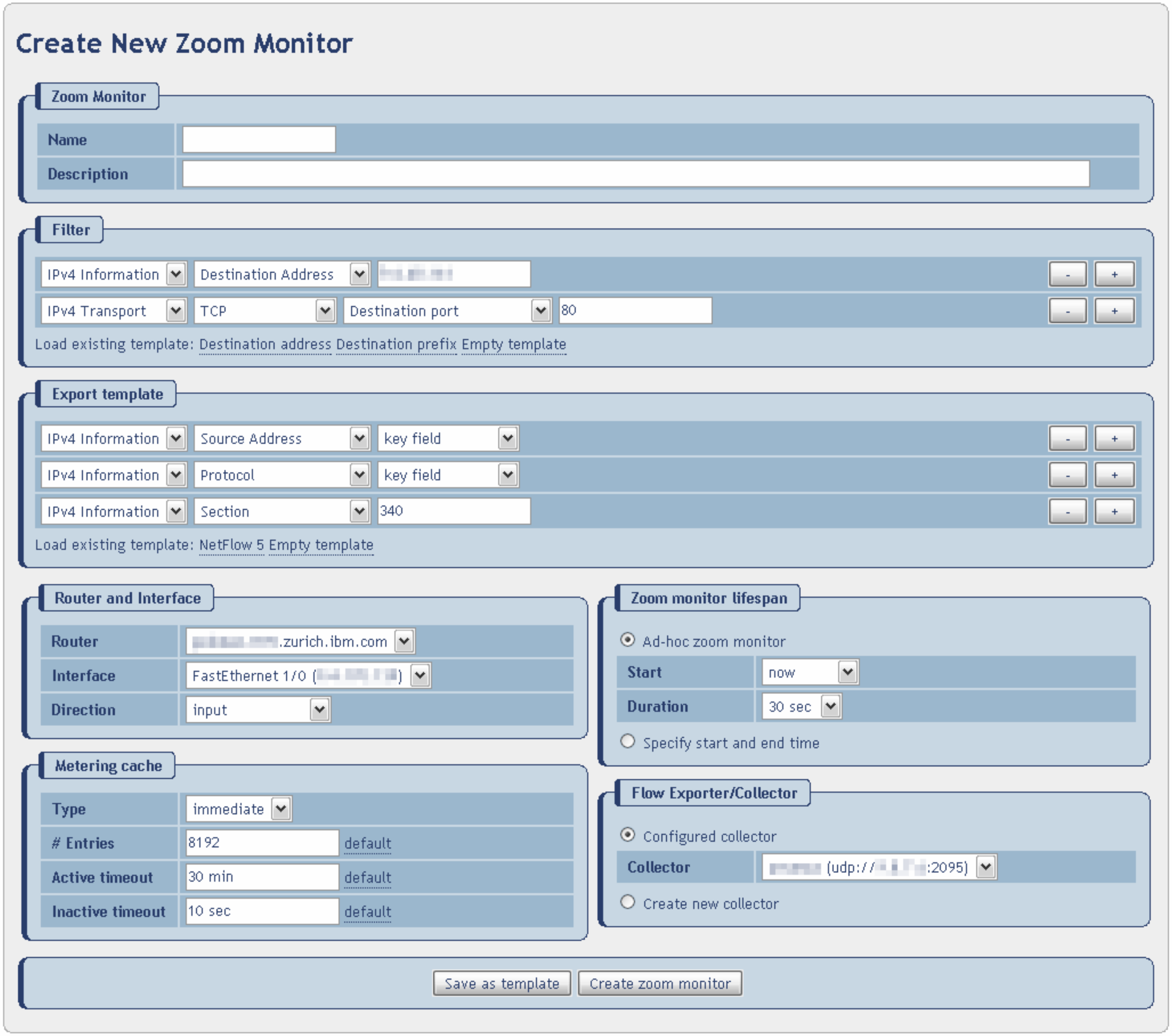
- **Technologies**
  - **Cisco IOS Flexible NetFlow (FNF)**
    - Configuration of multiple customized monitors
    - Currently: input filtering for FNF monitors not available (input filters needed at collector)
  - **Hespera Traffic Meter (IBM Research)**
    - Software-based flow monitor supporting NetFlow v5 and v9, IETF IPFIX exports
    - Customized flow exports (variable templates), CLI-based reconfiguration
    - Filtering with BPF filter syntax

- **User-based creation of dynamic zoom monitors**
  - Web-based specification of zoom monitors
  - Deployment on metering device (CLI-based) and management (e.g., lifespan)
    - Future: XML-based configuration (cf. [Dimitropoulos/Kind] or [NetConf])
  - Registering the zoom monitor at collector device (for disambiguation/triage)
  - Pre-defined zoom monitor templates from library

Export Collector

- **Prototype based on the Aurora flow analyzing system (IBM Research)**
  - Replaced existing Aggregation Database (ADB) with PostgreSQL (PG) backend
  - Input triage according to zoom monitors
  - Incremental population/gradually remove detailed representation: keep “Session”





Create New Zoom Monitor

Zoom Monitor

Name	
Description	

Filter

IPv4 Information ▾	Destination Address ▾	[redacted]		-	+
IPv4 Transport ▾	TCP ▾	Destination port ▾	80	-	+

Load existing template: Destination address Destination prefix Empty template

Export template

IPv4 Information ▾	Source Address ▾	key field ▾	-	+
IPv4 Information ▾	Protocol ▾	key field ▾	-	+
IPv4 Information ▾	Section ▾	340	-	+

Load existing template: NetFlow 5 Empty template

Router and Interface

Router	[redacted].zurich.ibm.com ▾
Interface	FastEthernet 1/0 ([redacted]) ▾
Direction	input ▾

Zoom monitor lifespan

● Ad-hoc zoom monitor

Start	now ▾
Duration	30 sec ▾

○ Specify start and end time

Metering cache

Type	immediate ▾	
# Entries	8192	default
Active timeout	30 min	default
Inactive timeout	10 sec	default

Flow Exporter/Collector

● Configured collector

Collector	[redacted] (udp://[redacted]:2095) ▾

○ Create new collector

Filter definition

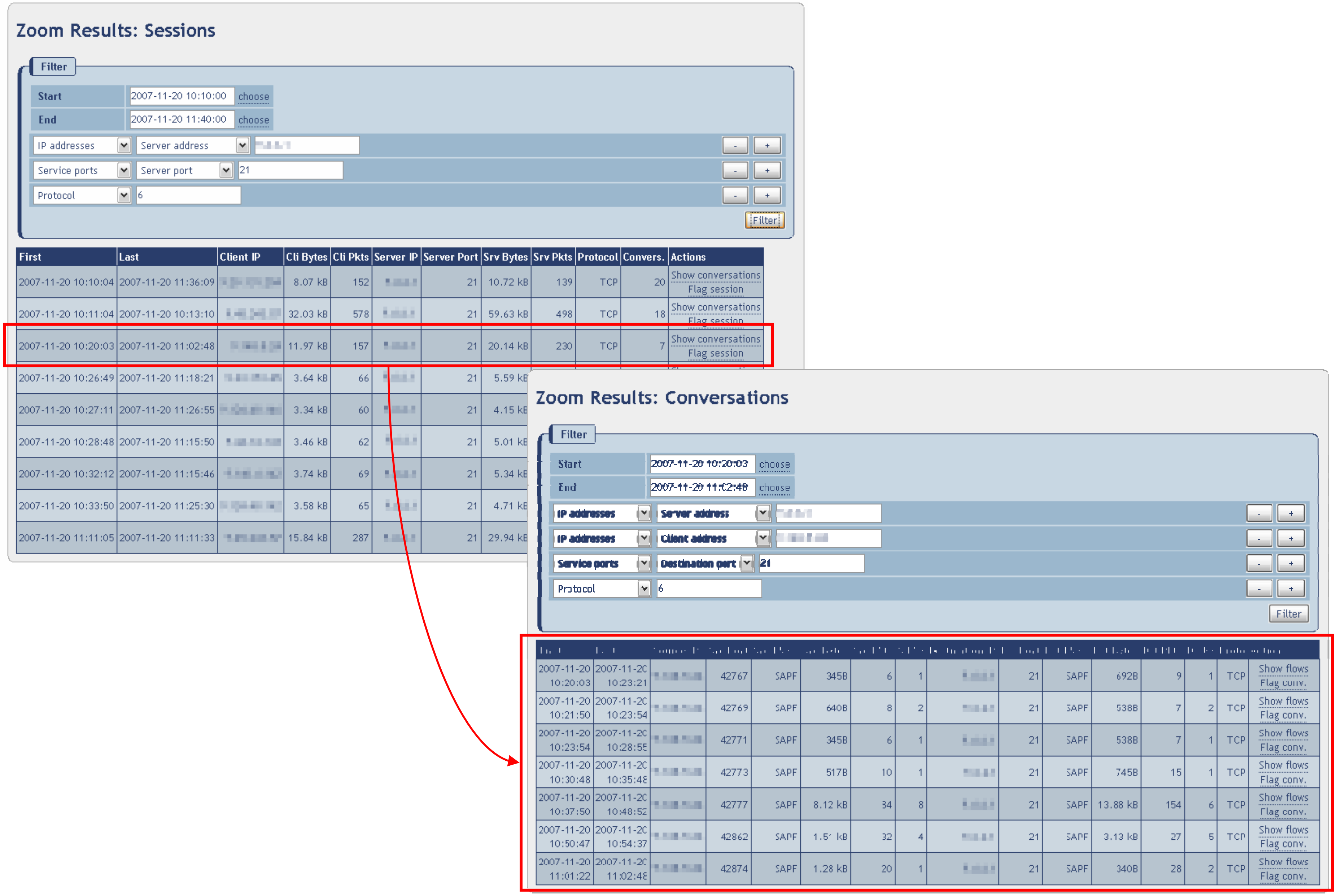
Export information

Router/Interface

Lifespan

Collector

Cache



Zoom Results: Sessions

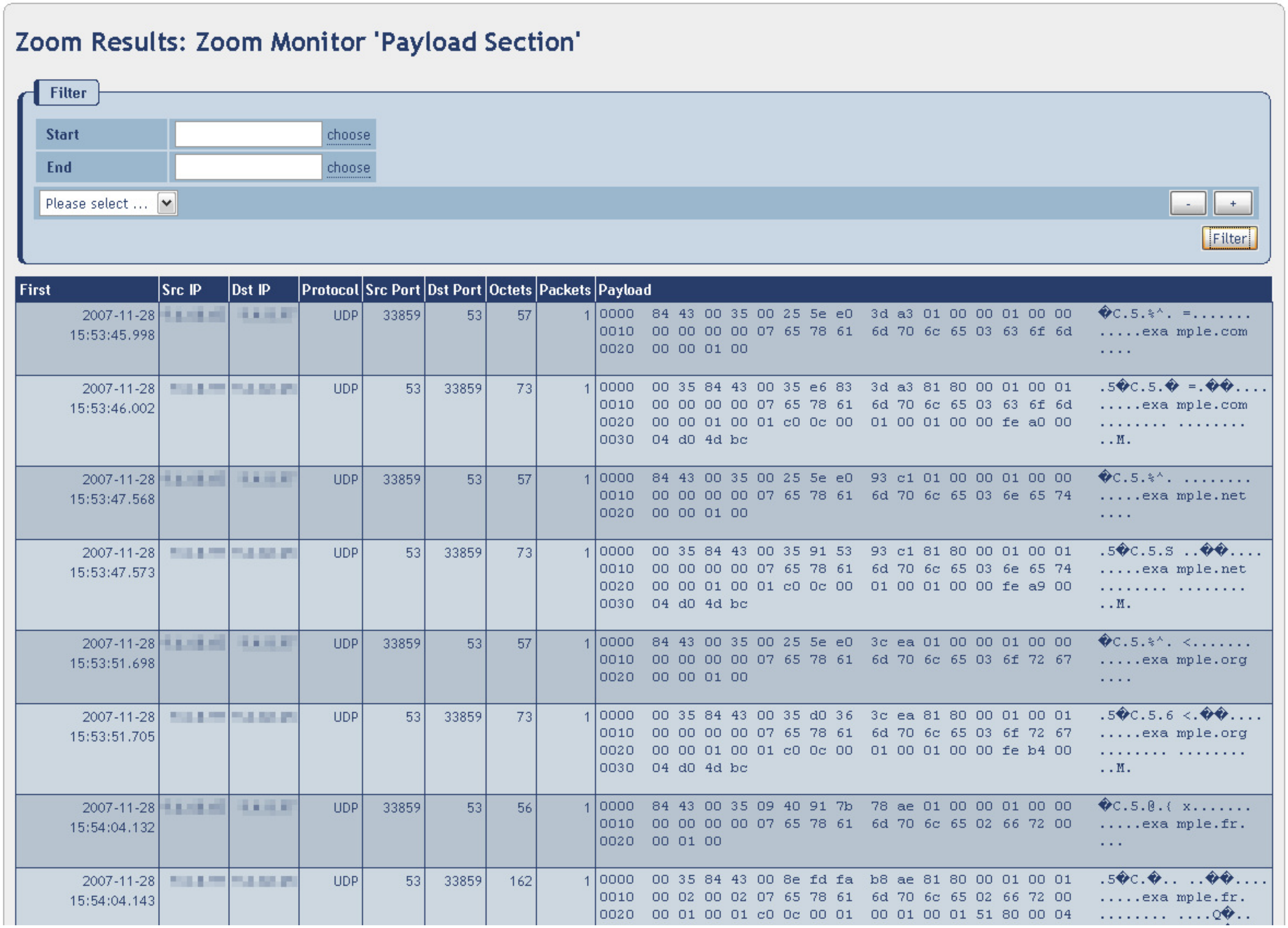
Start	2007-11-20 10:10:00 choose
End	2007-11-20 11:40:00 choose
IP addresses	Server address [redacted]
Service ports	Server port 21
Protocol	6

First	Last	Client IP	Cli Bytes	Cli Pkts	Server IP	Server Port	Srv Bytes	Srv Pkts	Protocol	Convers.	Actions
2007-11-20 10:10:04	2007-11-20 11:36:09	[redacted]	8.07 kB	152	[redacted]	21	10.72 kB	139	TCP	20	Show conversations Flag session
2007-11-20 10:11:04	2007-11-20 10:13:10	[redacted]	32.03 kB	578	[redacted]	21	59.63 kB	498	TCP	18	Show conversations Flag session
2007-11-20 10:20:03	2007-11-20 11:02:48	[redacted]	11.97 kB	157	[redacted]	21	20.14 kB	230	TCP	7	Show conversations Flag session
2007-11-20 10:26:49	2007-11-20 11:18:21	[redacted]	3.64 kB	66	[redacted]	21	5.59 kB				
2007-11-20 10:27:11	2007-11-20 11:26:55	[redacted]	3.34 kB	60	[redacted]	21	4.15 kB				
2007-11-20 10:28:48	2007-11-20 11:15:50	[redacted]	3.46 kB	62	[redacted]	21	5.01 kB				
2007-11-20 10:32:12	2007-11-20 11:15:46	[redacted]	3.74 kB	69	[redacted]	21	5.34 kB				
2007-11-20 10:33:50	2007-11-20 11:25:30	[redacted]	3.58 kB	65	[redacted]	21	4.71 kB				
2007-11-20 11:11:05	2007-11-20 11:11:33	[redacted]	15.84 kB	287	[redacted]	21	29.94 kB				

Zoom Results: Conversations

Start	2007-11-20 10:20:03 choose
End	2007-11-20 11:C2:48 choose
IP addresses	Server address [redacted]
IP addresses	Client address [redacted]
Service ports	Destination port 21
Protocol	6

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2007-11-20 10:20:03	2007-11-2C 10:23:21	[redacted]	42767	SAPF	345B	6	1	[redacted]	21	SAPF	692B	9	1	TCP	Show flows Flag conv.
2007-11-20 10:21:50	2007-11-2C 10:23:54	[redacted]	42769	SAPF	640B	8	2	[redacted]	21	SAPF	538B	7	2	TCP	Show flows Flag conv.
2007-11-20 10:23:54	2007-11-2C 10:28:55	[redacted]	42771	SAPF	345B	6	1	[redacted]	21	SAPF	538B	7	1	TCP	Show flows Flag conv.
2007-11-20 10:30:48	2007-11-2C 10:35:48	[redacted]	42773	SAPF	517B	10	1	[redacted]	21	SAPF	745B	15	1	TCP	Show flows Flag conv.
2007-11-20 10:37:50	2007-11-2C 10:48:52	[redacted]	42777	SAPF	8.12 kB	34	8	[redacted]	21	SAPF	13.88 kB	154	6	TCP	Show flows Flag conv.
2007-11-20 10:50:47	2007-11-2C 10:54:37	[redacted]	42862	SAPF	1.5' kB	32	4	[redacted]	21	SAPF	3.13 kB	27	5	TCP	Show flows Flag conv.
2007-11-20 11:01:22	2007-11-2C 11:02:48	[redacted]	42874	SAPF	1.28 kB	20	1	[redacted]	21	SAPF	340B	28	2	TCP	Show flows Flag conv.



Zoom Results: Zoom Monitor 'Payload Section'

First	Src IP	Dst IP	Protocol	Src Port	Dst Port	Octets	Packets	Payload

2007-11-28
15:53:45.998
	[redacted]	[redacted]	UDP	33859	53	57	1	`0000 84 43 00 35 00 25 5e e0 3d a3 01 00 00 01 00 00 �C.5.%^. =.......`
								`0010 00 00 00 00 07 65 78 61 6d 70 6c 65 03 63 6f 6d .....exa mple.com`
								`0020 00 00 01 00 ....`

2007-11-28
15:53:46.002
	[redacted]	[redacted]	UDP	53	33859	73	1	`0000 00 35 84 43 00 35 e6 83 3d a3 81 80 00 01 00 01 .5�C.5.� =.��....`
								`0010 00 00 00 00 07 65 78 61 6d 70 6c 65 03 63 6f 6d .....exa mple.com`
								`0020 00 00 01 00 01 c0 0c 00 01 00 01 00 00 fe a0 00 ........ ........`
								`0030 04 d0 4d bc ..M.`

2007-11-28
15:53:47.568
	[redacted]	[redacted]	UDP	33859	53	57	1	`0000 84 43 00 35 00 25 5e e0 93 c1 01 00 00 01 00 00 �C.5.%^. ........`
								`0010 00 00 00 00 07 65 78 61 6d 70 6c 65 03 6e 65 74 .....exa mple.net`
								`0020 00 00 01 00 ....`

2007-11-28
15:53:47.573
	[redacted]	[redacted]	UDP	53	33859	73	1	`0000 00 35 84 43 00 35 91 53 93 c1 81 80 00 01 00 01 .5�C.5.S ..��....`
								`0010 00 00 00 00 07 65 78 61 6d 70 6c 65 03 6e 65 74 .....exa mple.net`
								`0020 00 00 01 00 01 c0 0c 00 01 00 01 00 00 fe a9 00 ........ ........`
								`0030 04 d0 4d bc ..M.`

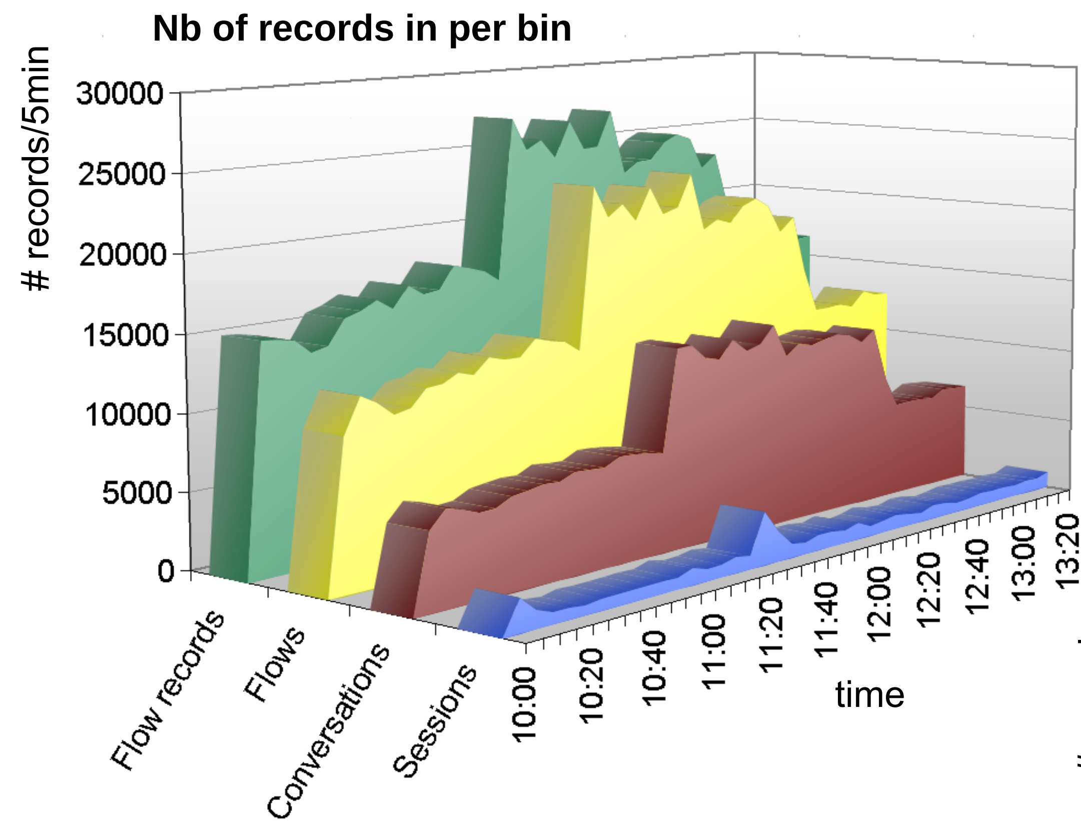
2007-11-28
15:53:51.698
	[redacted]	[redacted]	UDP	33859	53	57	1	`0000 84 43 00 35 00 25 5e e0 3c ea 01 00 00 01 00 00 �C.5.%^. <.......`
								`0010 00 00 00 00 07 65 78 61 6d 70 6c 65 03 6f 72 67 .....exa mple.org`
								`0020 00 00 01 00 ....`

2007-11-28
15:53:51.705
	[redacted]	[redacted]	UDP	53	33859	73	1	`0000 00 35 84 43 00 35 d0 36 3c ea 81 80 00 01 00 01 .5�C.5.6 <.��....`
								`0010 00 00 00 00 07 65 78 61 6d 70 6c 65 03 6f 72 67 .....exa mple.org`
								`0020 00 00 01 00 01 c0 0c 00 01 00 01 00 00 fe b4 00 ........ ........`
								`0030 04 d0 4d bc ..M.`

2007-11-28
15:54:04.132
	[redacted]	[redacted]	UDP	33859	53	56	1	`0000 84 43 00 35 09 40 91 7b 78 ae 01 00 00 01 00 00 �C.5.@.{ x.......`
								`0010 00 00 00 00 07 65 78 61 6d 70 6c 65 02 66 72 00 .....exa mple.fr.`
								`0020 00 01 00 ...`

2007-11-28
15:54:04.143
	[redacted]	[redacted]	UDP	53	33859	162	1	`0000 00 35 84 43 00 8e fd fa b8 ae 81 80 00 01 00 01 .5�C.�.. ..��....`
								`0010 00 02 00 02 07 65 78 61 6d 70 6c 65 02 66 72 00 .....exa mple.fr.`
								`0020 00 01 00 01 c0 0c 00 01 00 01 00 01 51 80 00 04 ........ ....Q�..`

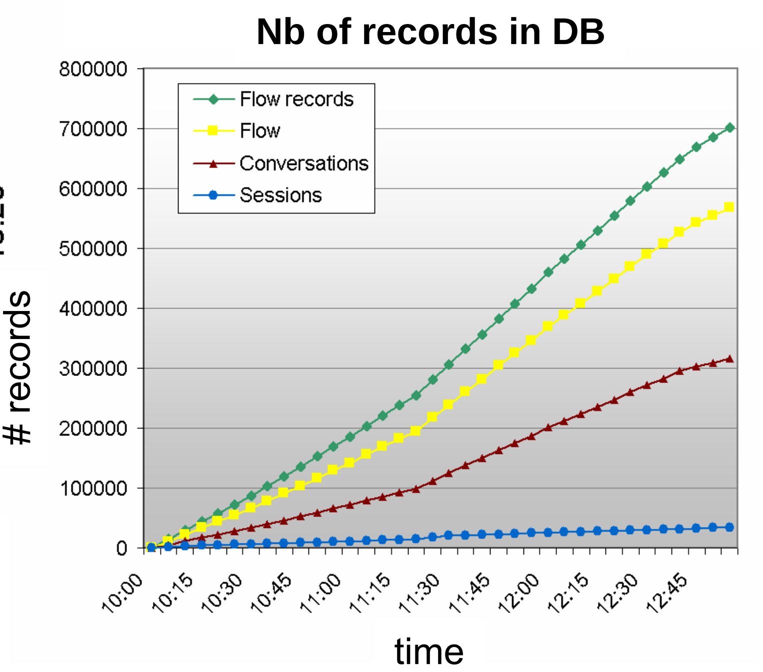
Results: Compression (WAN traffic)



Nb of records in per bin

- **Average compression ratio**

#flow records : #flows	**1.26**	σ = 0.07
#flow records : #conversations	**2.34**	σ = 0.28
#flow records : #sessions	**22.80**	σ = 7.00

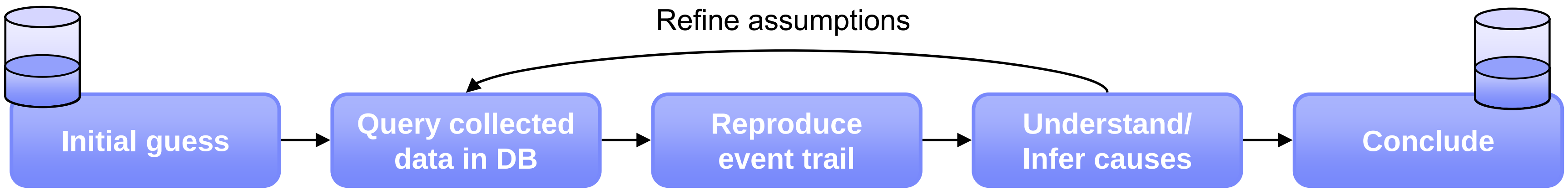


Nb of records in DB

- Session inactive timeout: 20min

Traffic Collection for Incident Analysis

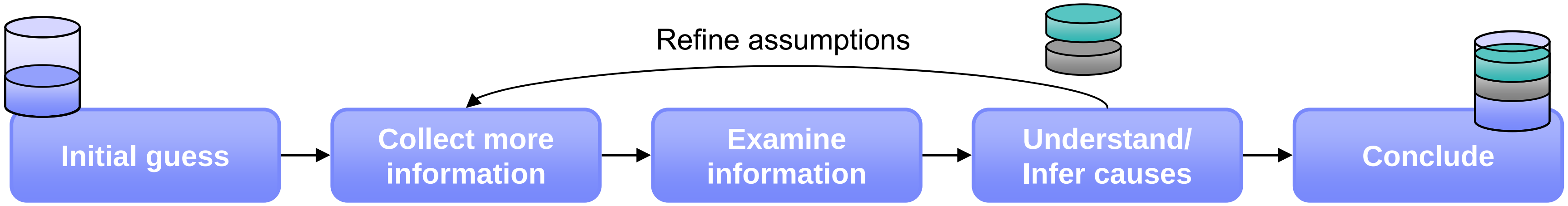
- **After-the-fact analysis**



Refine assumptions

Initial guess → Query collected data in DB → Reproduce event trail → Understand/ Infer causes → Conclude

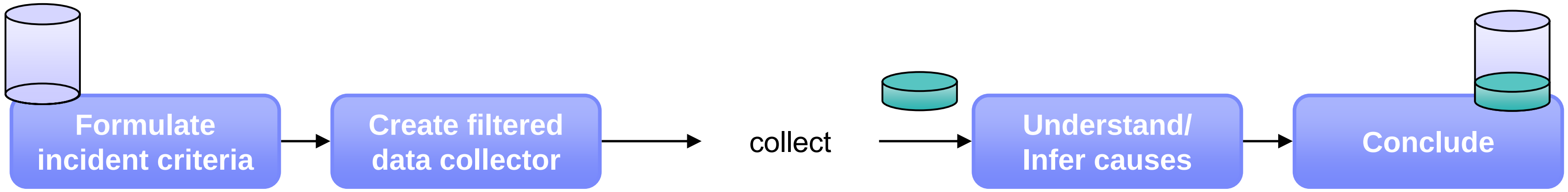
- **Real-time analysis**



Refine assumptions

Initial guess → Collect more information → Examine information → Understand/ Infer causes → Conclude

- **Future incident trap**



Formulate incident criteria → Create filtered data collector → collect → Understand/ Infer causes → Conclude

Future Work and Vision

- **Automated zoom monitor creation**
  - Interface to a behavior-based network anomaly detection system
  - Proactive collection of evidence for off-line forensic analysis of abnormal events
- **Distributed collector infrastructure**
  - Distributed collectors, e.g., at multiple sites (scalability)
  - Transfer required information to central reporting system on demand
- **Cisco IOS Flexible NetFlow with input filters**
  - Perform filtering on routers to replace software-based metering (and filtering)

Conclusion

- **Incident analysis tool adapting flow information granularity**
  - Increase level of detail of relevant/unknown traffic events
  - Decrease level of detail (lossy compression) of less relevant events
  - Keep a meaningful abstraction of all traffic events

- **Creation of customized zoom monitors**
  - Zoom in on specific traffic to gain additional information about its properties and behavior
  - Centralized management of metering devices for traffic detail collection

References

- IBM Research. “Aurora – Network Traffic Analysis and Visualization”. http://www.zurich.ibm.com/aurora/
- Xenofontas Dimitropoulos and Andreas Kind. “Configuration of Monitors”. FloCon2008.
- NETCONF IETF Working Group. http://www.ops.ietf.org/netconf/
- Cisco Systems, Inc. “Cisco IOS Flexible Netflow”. Product website: http://www.cisco.com/en/US/products/ps6965/products_ios_protocol_option_home.html