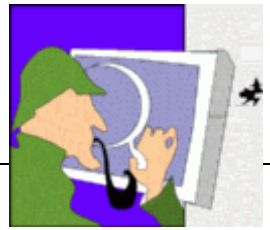
Case Study:
Instrumenting a Network for NetFlow Security Visualization Tools

William Yurcik* **Yifan Li**

SIFT Research Group

National Center for Supercomputing Applications (NCSA)

University of Illinois at Urbana-Champaign



Annual Computer Security Applications Conference (ACSAC'05)

Overview

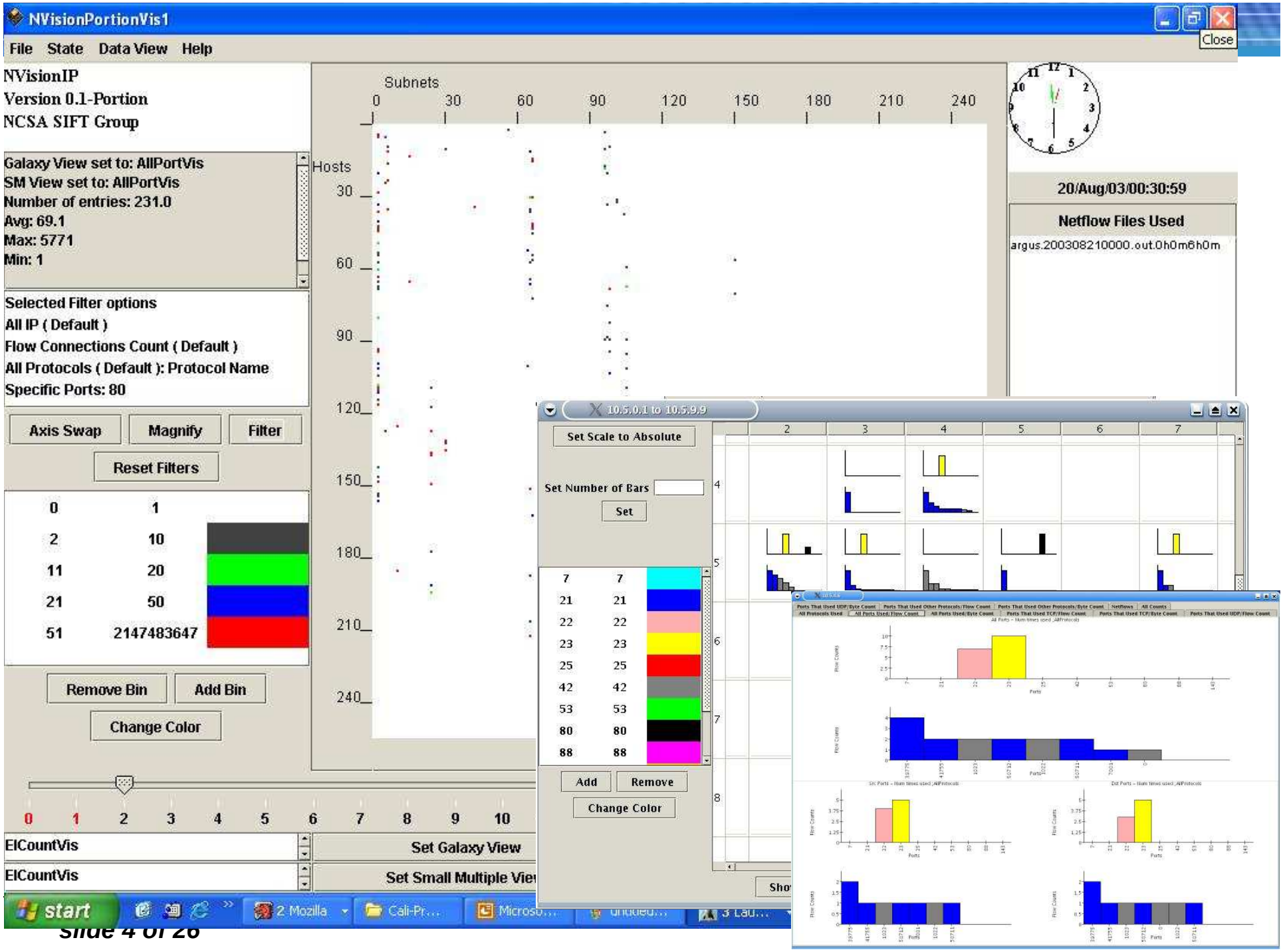
- NetFlows for Security?
- What are NetFlows?
- Challenges with NetFlows
  - Instrumentation
  - Data Management
  - Interoperability
  - Sampling
- Conclusions

NCSA

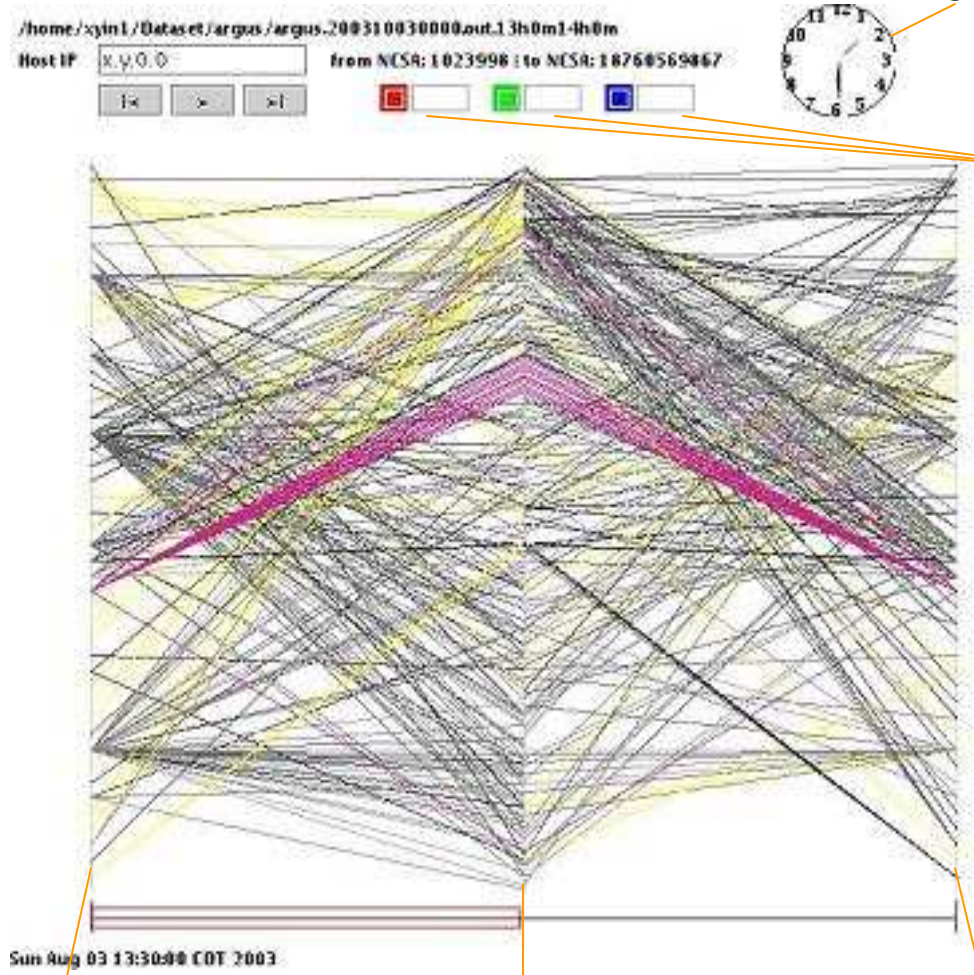
NetFlow Security Analysis

- Who are my top N talkers
  - What percentage of traffic are they?
- How many users are on the network at any given time?
  - When will upgrades effect the least number of users?
- How long do my users surf?
- Where do they go?
- Where did they come from?
- Are users following the security policy?
- Alarm connection-oriented attacks like DoS, DDoS, malware distrribution, worm scanning, vulnerability scanning, etc…
  - Will watch for these attacks destined for anywhere or coming from anywhere or entirely within your perimeter!

NCSA


NVisionPortionVis1
File State Data View Help
NVisionIP
Version 0.1-Portion
NCSA SIFT Group
Galaxy View set to: AllPortVis
SM View set to: AllPortVis
Number of entries: 231.0
Avg: 69.1
Max: 5771
Min: 1
Selected Filter options
All IP (Default)
Flow Connections Count (Default)
All Protocols (Default): Protocol Name
Specific Ports: 80
Axis Swap
Magnify
Filter
Reset Filters
Remove Bin
Add Bin
Change Color
Subnets
Hosts
20/Aug/03/00:30:59
Netflow Files Used
argus.200308210000.out.0h0m6h0m
10.5.0.1 to 10.5.9.9
Set Scale to Absolute
Set Number of Bars
Set
Add
Remove
Change Color
ElCountVis
Set Galaxy View
Set Small Multiple Vie
start
2 Mozilla
Cali-Pr...
Microso...
untitled...
3 Lau...

VisFlowConnect External View



clock

highlighted ports

time axis

outside domains

inside hosts

outside domains

What are NetFlows?

What NetFlows are…

- Information about packets sent between two network nodes
  - source and destination IP addresses
  - source and destination TCP/UDP ports
  - IP protocol type
  - number of packets and bytes (octets)
  - start and stop time

NetFlow Data Record

Usage
- Packet Count
- Byte Count

Time of Day
- Start Timestamp
- End Timestamp

Port Utilization
- Input Interface Port
- Output Interface Port

QoS
- Type of Service
- TCP Flags
- Protocol

From/To
- Source IP Address
- Destination IP Address

Application
- Source TCP/UDP Port
- Destination TCP/UDP Port

Routing and Peering
- Next Hop Address
- Source AS Number
- Dest. AS Number
- Source Prefix Mask
- Dest. Prefix Mask

NCSA

Why NetFlows?

Data Source	Description	Advantage	Disadvantage
Packet	lowest level of granularity; all raw packets with all fields intact	most detailed data and statistics especially protocols; easiest to obtain	unscalable; protocol signaling needs to be decoded
NetFlows	IPs/ports/protocols/ Timestamps/data?	scalable for catching all traffic; multiple sources, uniform field formats	maybe no data field; context must be inferred
IDS	alerts of different formats	scalable; tunable	resource-intensive; misses; FPs
Load Levels	aggregate utilization levels that can be broken down to IP, protocol, port	high volume attacks (DOS, traffic); capacity planning; availability from routers & sniffers	details about SD pairs; no direction; low volume events obscured

NCSA

What NetFlows are not…

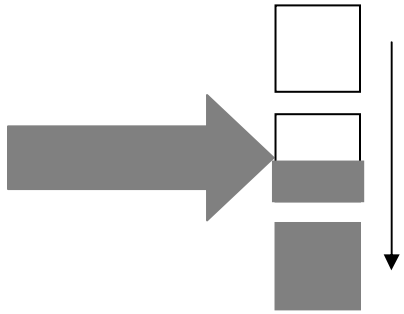
- Infinitely Scalable Data Source
  - Unreliable UDP network transfer for Cisco
  - Argus reliability dependent on CPU speed and I/O of sensor PC
- Content aware
  - i.e., flow data may not contain packet payload, just metadata about the packets
- Proactive/Preventative Security Sensor
  - near-real time is good enough
  - layered defense still required – black-listing must be coordinated with other security devices

NCSA

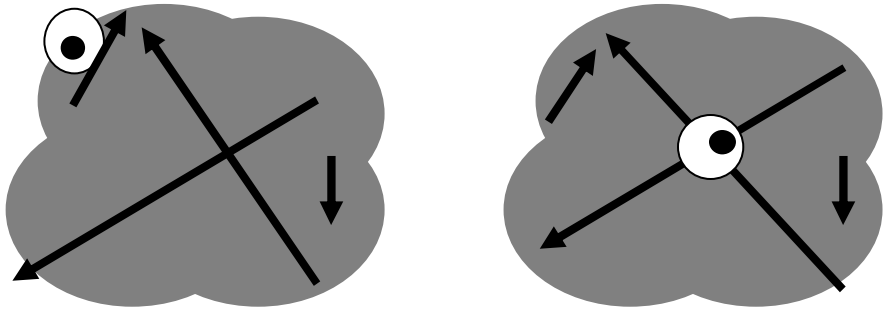
NetFlows Network Instrumentation Issues

NetFlows Instrumentation Issues

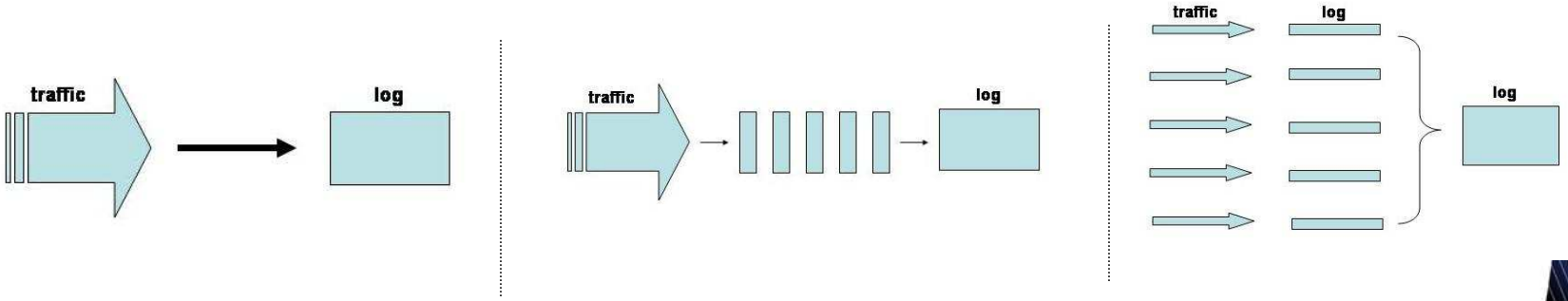
- Streaming Data



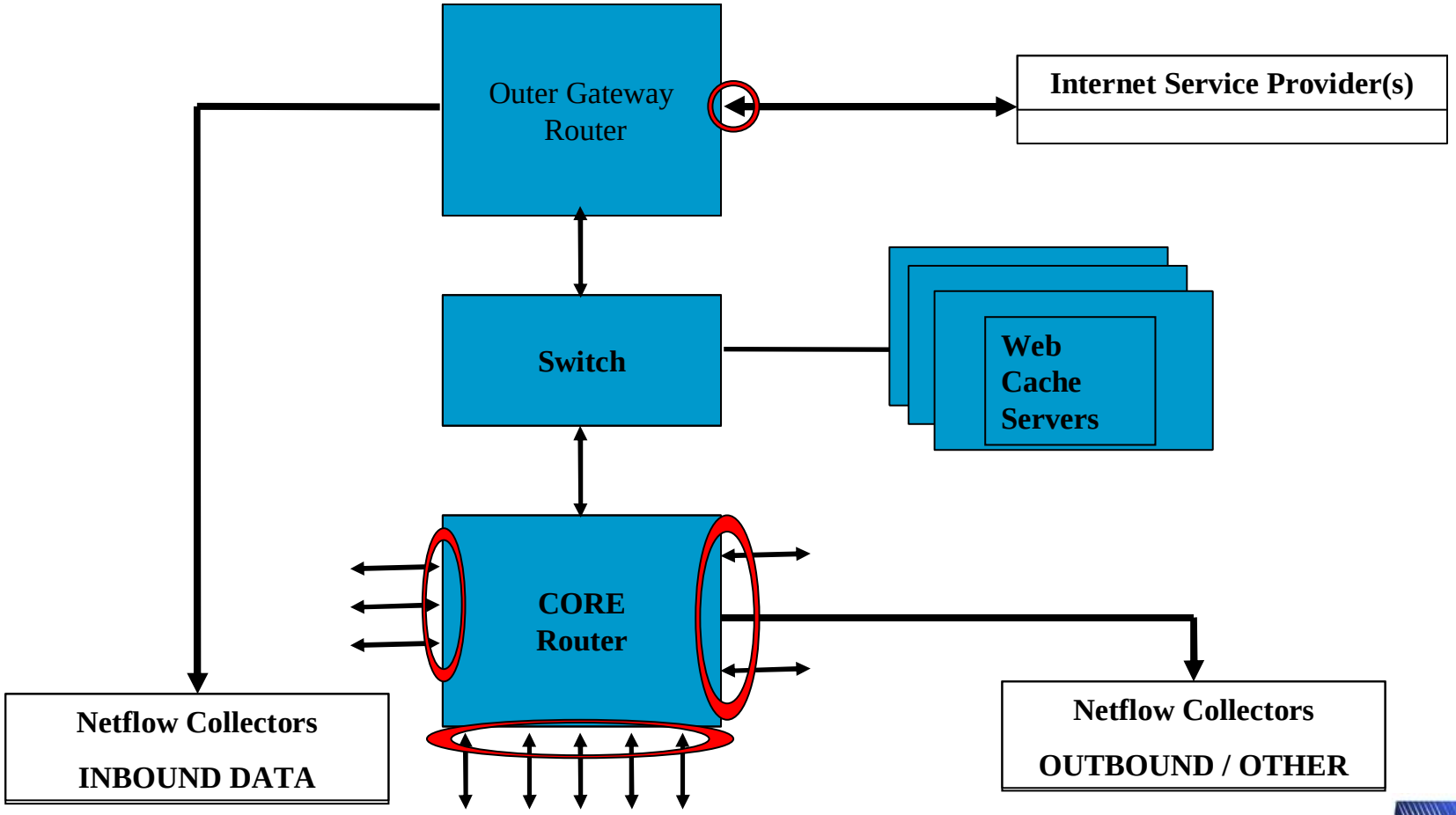
- Vantage Point



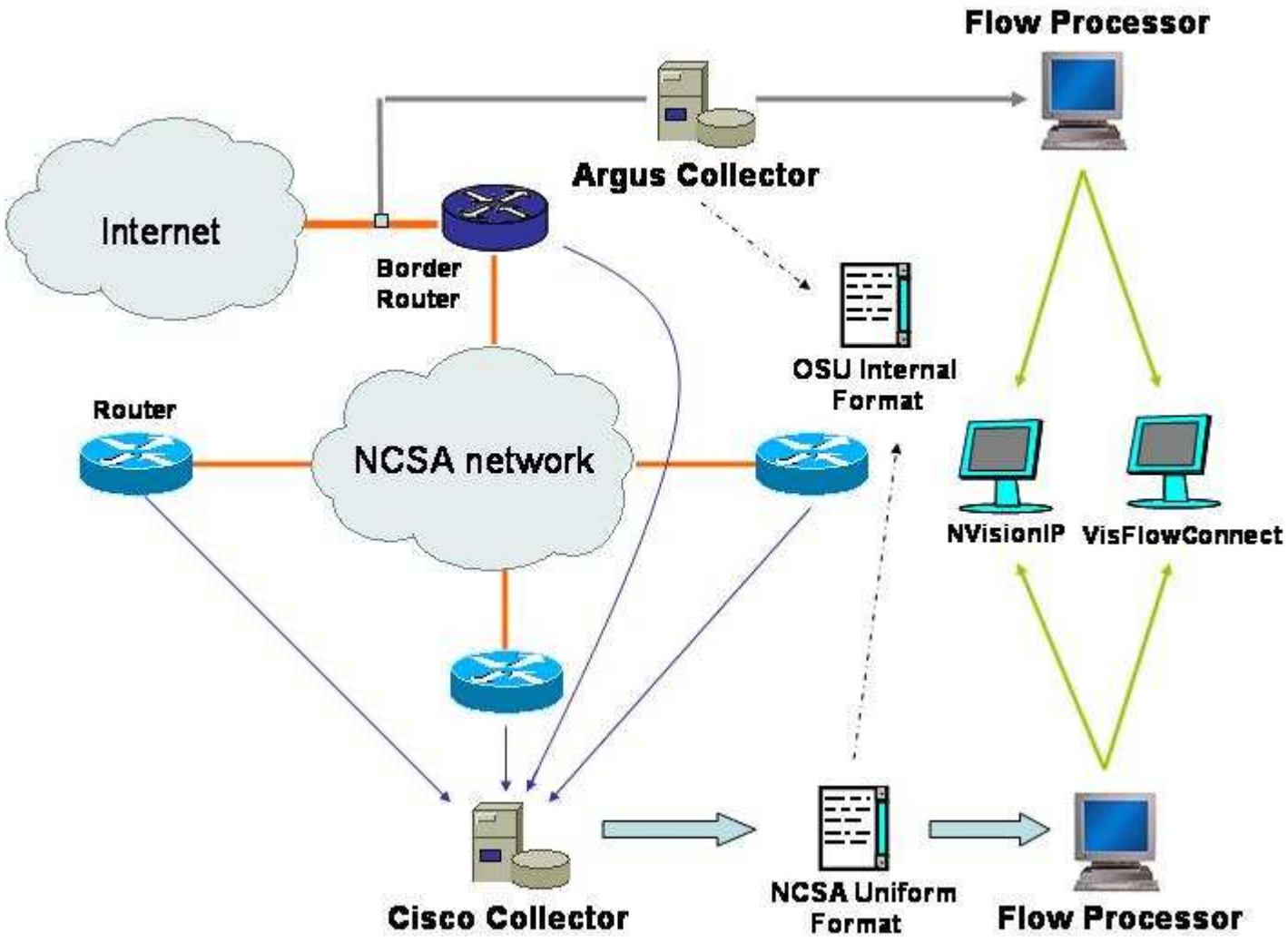
- High Line Rates



NetFlow Collection Points

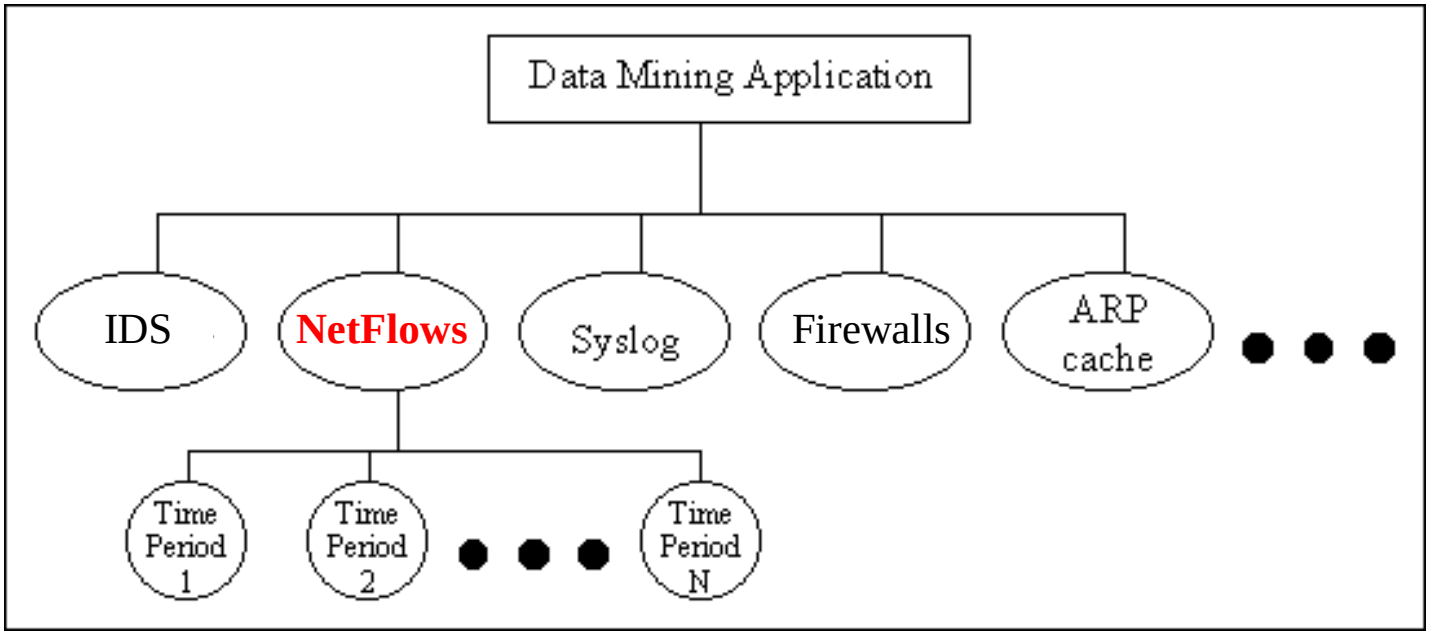


NCSA's NetFlows Architecture



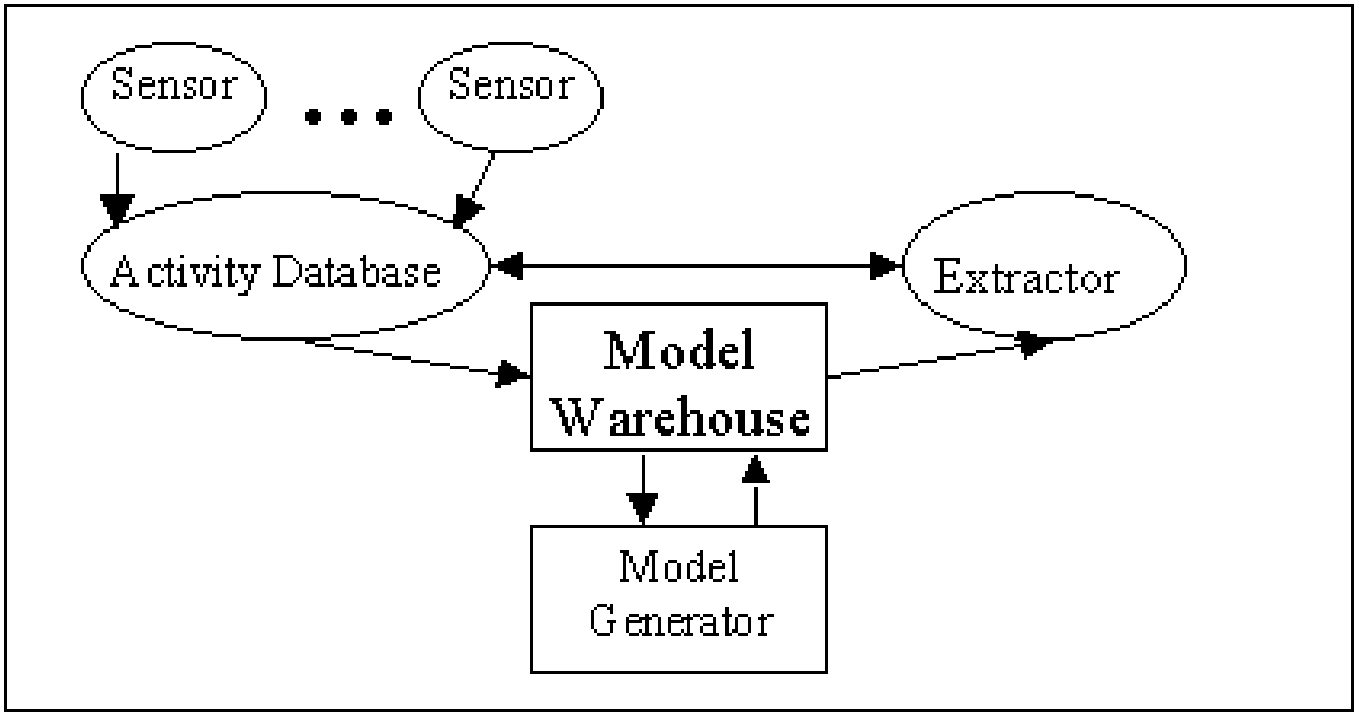
NetFlows Data Management Issues

The NetFlows Data Management Problem



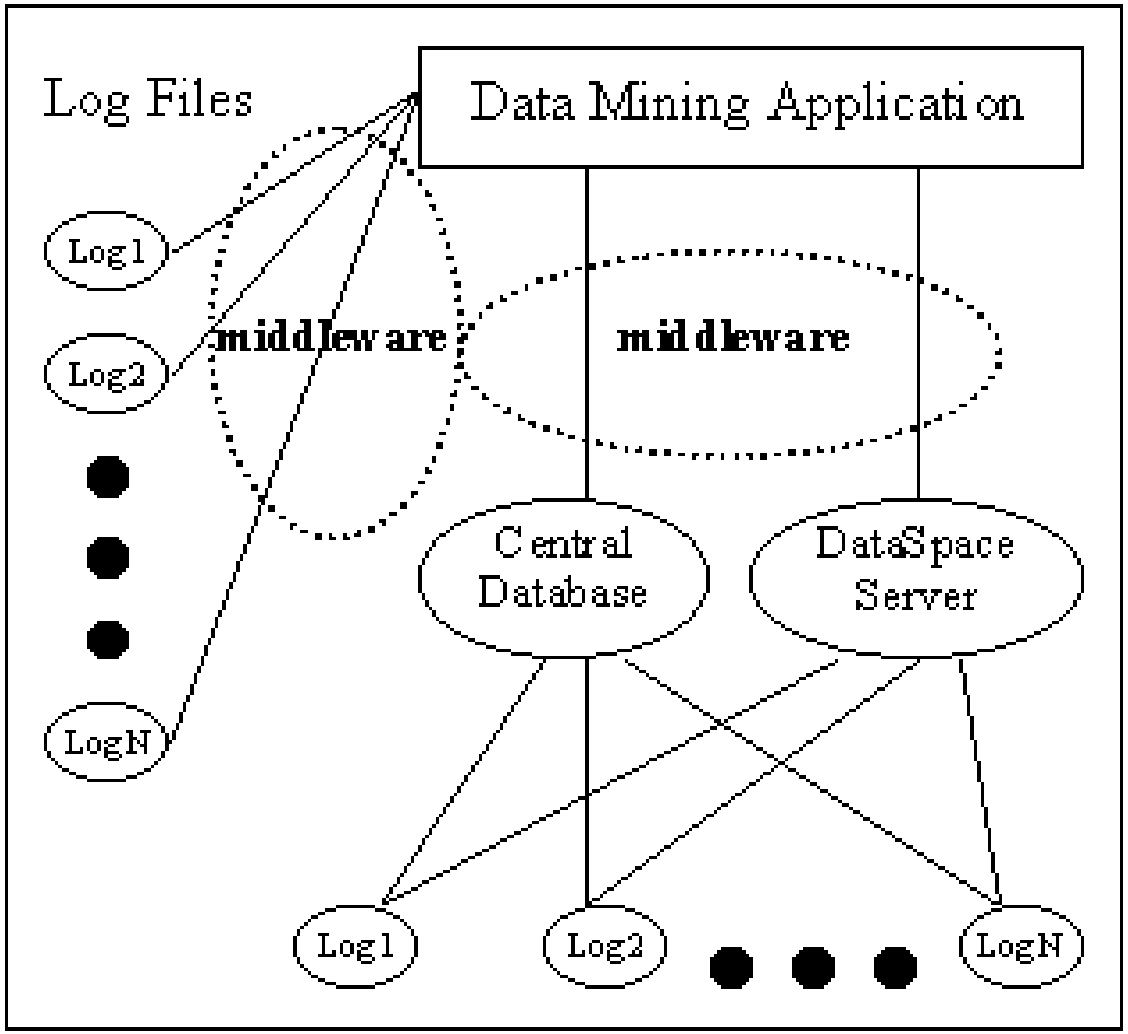
time dimension

NetFlows Data Management: Alternative 1



(1) Central Database Architecture

NetFlows Data Mananagement: Alternative 2



(2) Middleware Architecture

Data Management Problem Summary

- A simple matter of economics:
  - Database storage = LARGE investment
    - A larger than 10 TB database can cost tens of millions of dollars (particularly in 2001)
    - Although increasing there are finite limits even with unlimited budget.
  - Data storage in flat files
    - No need to move large volume raw data sources.
    - Multiple machine support.
    - Backup in place, oldest data to tape.
    - Large SANs can be used to store data reliably and are relatively scalable, but are still expensive.
- Are there alternatives to these two basic data storage alternatives? *research needed here*

NCSA

NetFlows Interoperability Issues

Interoperability: Flow Flavors

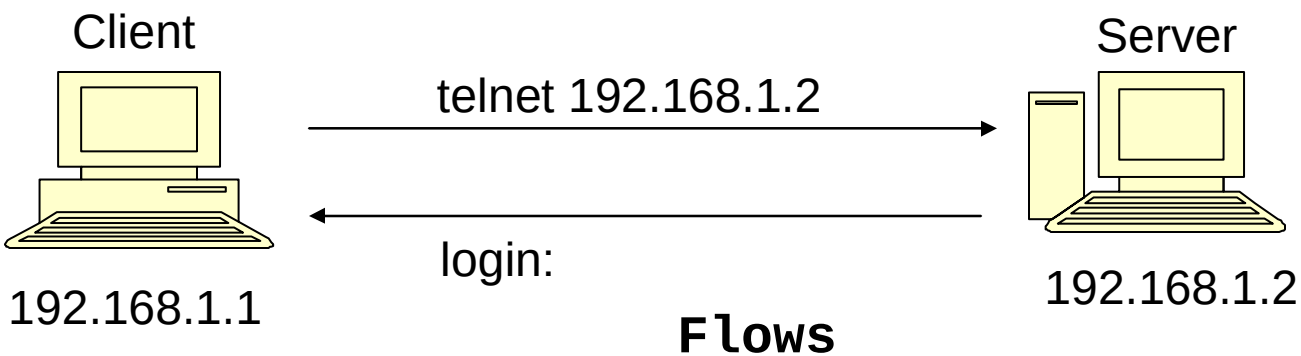
- **Router-Based NetFlows (Cisco, Juniper, Foundry)**
  - **unidirectional** →
  - **commercial**
  - **cache timeout (30 min)**
  - **configuration - sampling**

- **Platform-Independent NetFlows (Argus)**
  <http://www.qosient.com/argus/>
  - **bidirectional** ↔
  - **open source**
  - **cache timeout (1 min)**
  - **configuration – collect data field**

- **Other**
  - **sFlows, NFDump, etc.**

NCSA

Directionality

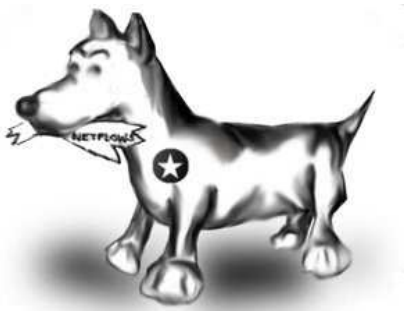


Flows

SrcIP	SrcPort	DestIP	DestPort	Protocol	
192.168.1.1	1234	192.168.1.2	23	TCP	x bytes
192.168.1.2	23	192.168.1.1	1234	TCP	x bytes

192.168.1.2	23	192.168.1.1	1234	TCP	x bytes

Interoperability Solution: CANINE Tool



- **C**onverter and **AN**onymizer for **I**nvestigating **N**etflow **E**vents
- Handles several NetFlow formats
  - Cisco V5 & V7, ArgusNCSA, CiscoNCSA, NFDump
- Also provides multi-dimensional anonymization of fields to facilitate secure data sharing
- URL:
  <http://security.ncsa.uiuc.edu/distribution/CanineDownLoad.html>

NCSA

NetFlows Sampling Issues

NetFlows Sampling

- Systematic Sampling (deterministic function)
  - Count-based (spatial packet position; e.g., packet count)
  - Time-based (temporal packet position; e.g., arrival time)
- Random Sampling
  - n-out-of-N
  - Probabilistic
    - Uniform Probabilistic (same probability for each packet)
    - Non-Uniform Probabilistic (probability depends on input)
    - Flow State Probabilistic
  - Sampling probability depends on flow state

NCSA

Summary and Future Work

- **NetFlows network instrumentation for security is specialized and non-trivial, however, the payoff is large**
- **Flow-Analysis Community Homepage**

<http://www.ncassr.org/projects/sift/flow-analysis/>

Future Work:

NetFlows anonymization to facilitate security data-sharing

NCSA SIFT Project

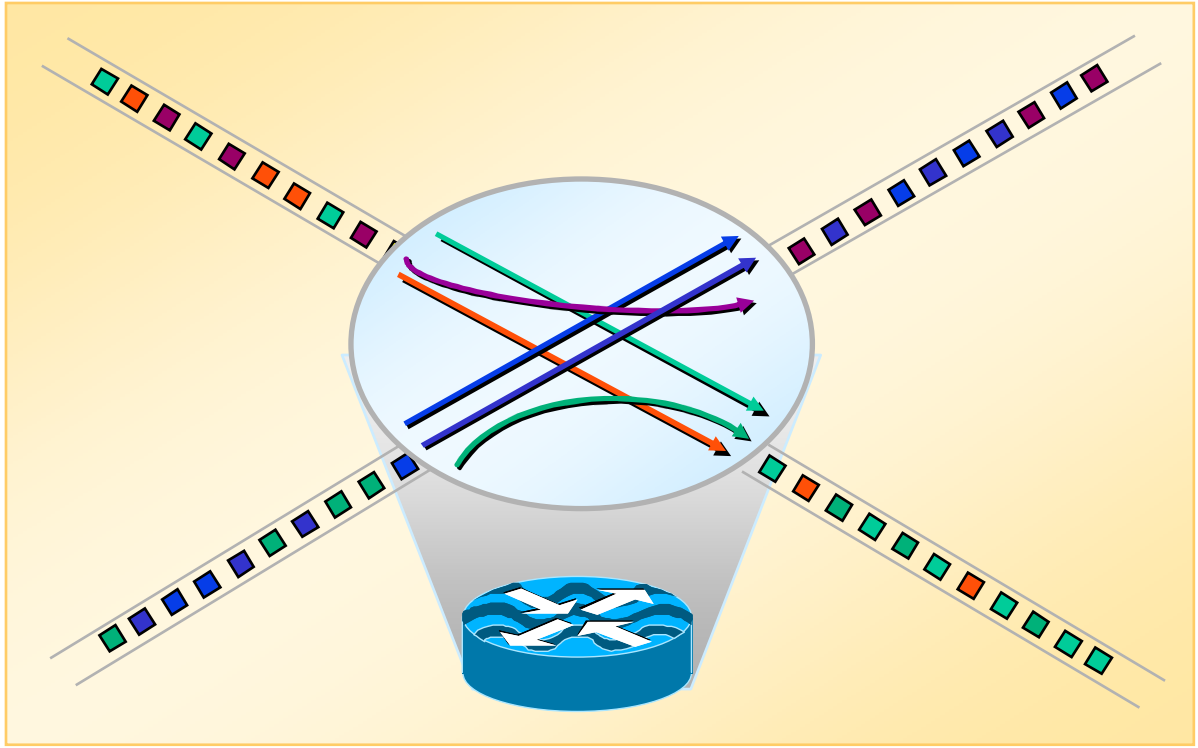
<http://www.ncassr.org/projects/sift/>

Questions?

Flow-Based Analysis

Seven Keys Define a Flow:

1. Source Address
2. Destination Address
3. Source Port
4. Destination Port
5. Layer 3 Protocol
6. TOS Byte (DSCP)
7. Input Interface



Project Motivations

- **NetFlows in multiple, incompatible formats**
  - **Network security monitoring tools usually support one or two NetFlows format**
  - **Need conversion of NetFlows between different formats**
- **Sensitive network information hinders log sharing**
  - **Log sharing necessary for research and study**
  - **Need anonymization of sensitive data fields**

NCSA