The CINBAD Project

24 April 2008

Dan Ford – HP/ProCurve
with contributions (alphabetical):
Milosz Hulboj - CERN/Procurve
Jean-Michel Jouanigot - CERN
Ryszard Jurga - CERN/Procurve
Arnaud Pierson – HP Labs

Outline

- Project description
- Accomplishments (July ’07—April ’08)
- Next steps

Project Description

Codename: “CINBAD”

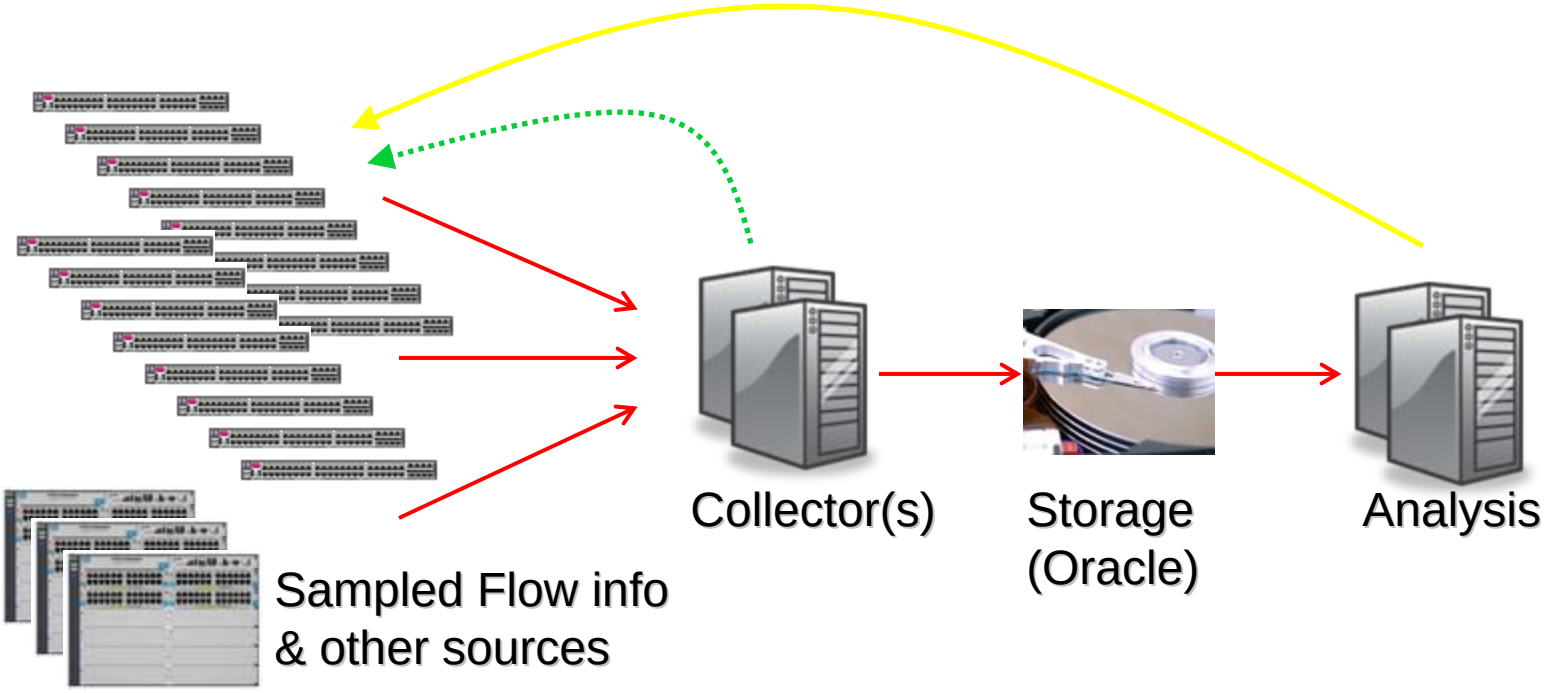
CERN **I**nvestigation of **N**etwork **B**ehaviour and **A**nomaly **D**etection

Project Goal

“T*o **understand the behaviour of large computer networks** (10’000+ nodes) in High Performance Computing or large Campus installations to be able to:*

- *Detect traffic anomalies in the system*
- *Be able to perform trend analysis*
- *Automatically take counter measures*
- *Provide post-mortem analysis facilities* “

Basic Principle



Investigation topics

- Challenging research activity
- Must address large scale issues
  - Collection of large quantity of data
  - Storage & post mortem
  - Analysis
- Requires initial understanding of precise definitions and heuristics
  - Anomalies?
  - Trends?
  - Counter measures…

Deliverables

- The project is tentatively divided into three phases, each with a particular set of investigation activities and deliverables
  - **Data collection and network management**
  - **Data Analysis and algorithm development**
  - **Performance and scalability analysis**

Activities and Accomplishments

(July 2007—April 2008)

Achievements

Packet Sampling Studies

- Over 100 technical papers read and analysed
- Thorough Technical Report available on the CINBAD website
  (http://openlab-mu-internal.web.cern.ch/openlab-mu-internal/openlab-II_Projects/SamplingReport.pdf)
- Indicative of relevance and potential of this area of investigation
  - Few studies that specifically addressed sFlow and anomaly detection

Understanding the data sources (sflow)

- In-depth analysis of sflow agents and their performance (switches)
- Simulation of sampling
  - Full data collected on live network
  - Various sampling algorithms simulated and evaluated
  - -> Ways to estimate real traffic from samples
  - -> Measure of the quality of the estimation
- Established contact with HP Labs and Berkeley to improve the above

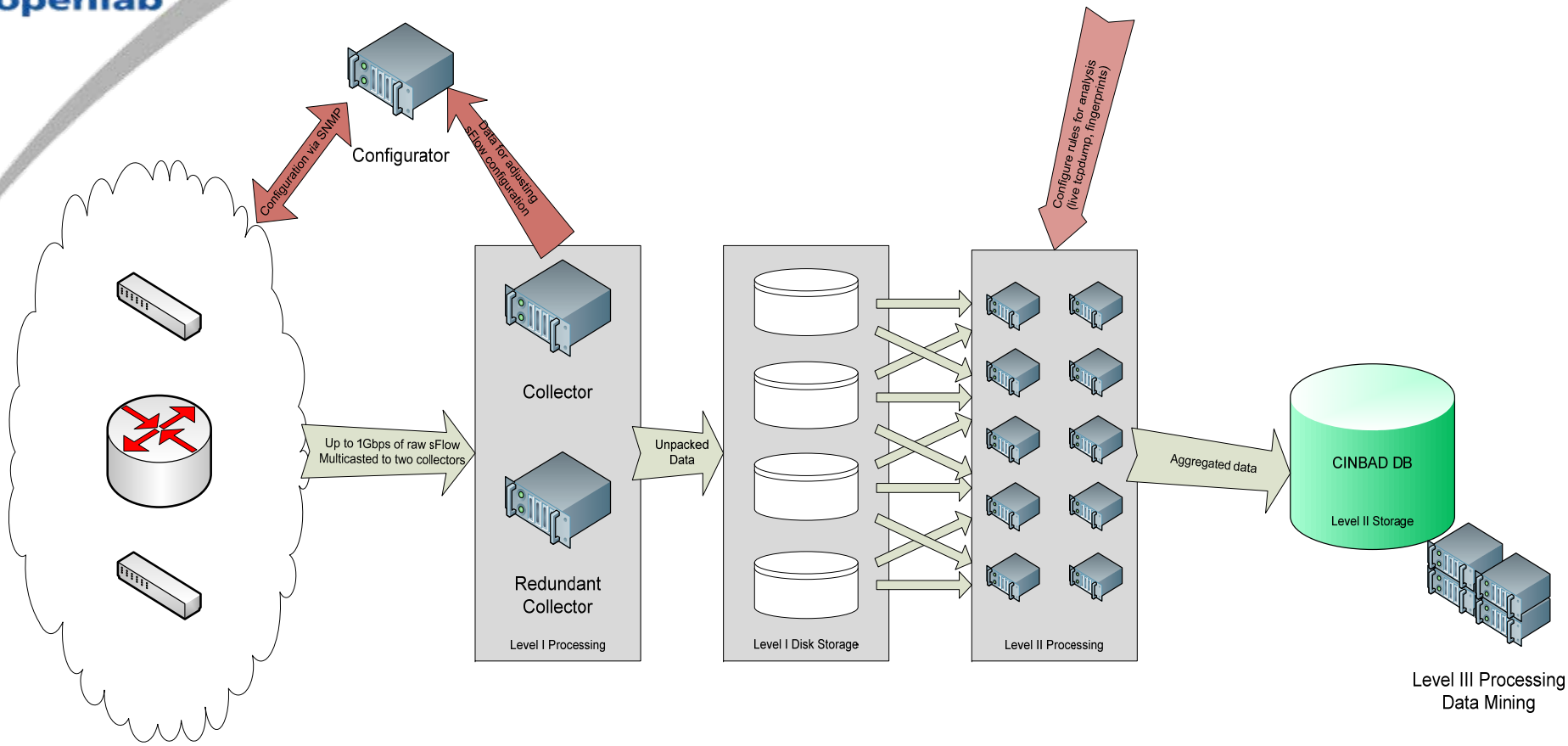
What is an anomaly?

- Many people consulted to understand their wishes
  - CERN network team
  - CERN security team
  - IDM @ Procurve
- Common patterns identified
  - Rogue network service detection
  - Malicious traffic
  - Post mortem incident correlation

Survey on data acquisition @ CERN

- Estimated data collected: 300’000 samples/s
- Survey of
  - Current Oracle and application performance in use at CERN: Lemon, PVSS, etc
  - LHC experiments experts consulted:
    - High performance Data storage
    - Data format and representation
    - Analysis principles
- Conclusion: follow a two level strategy

Collector Design



Highly Scalable Architecture

Rich database for investigative data mining

Next Steps

Collect sizable data

- Implement one high performance collector and storage
- Implement a configuration mechanism to setup the agents
- Collect sflow samples from all CERN devices for several days
- Identify potential performance issues

Data mining

- Analyse the data stored:
  - Define the characteristics of “zero-day” anomalies
  - Understand the possible correlation on the data and other sources (antivirus, intrusion detection systems, network incidents, layer 3/route changes, etc)
  - Identify the best candidates for database storage and further analysis on historical data

Conclusion

- This research activity has created a strong interest within HP/Procurve and CERN
  - Very open and friendly network established
- An in-depth compilation of the work done in packet sampling techniques is complete
- We achieved the design of a scalable collector
- Looking forward to a continuing fruitful collaboration
  - Fellows will be visiting ProCurve again in May