CINBAD

Major Review Meeting

29 September 2009

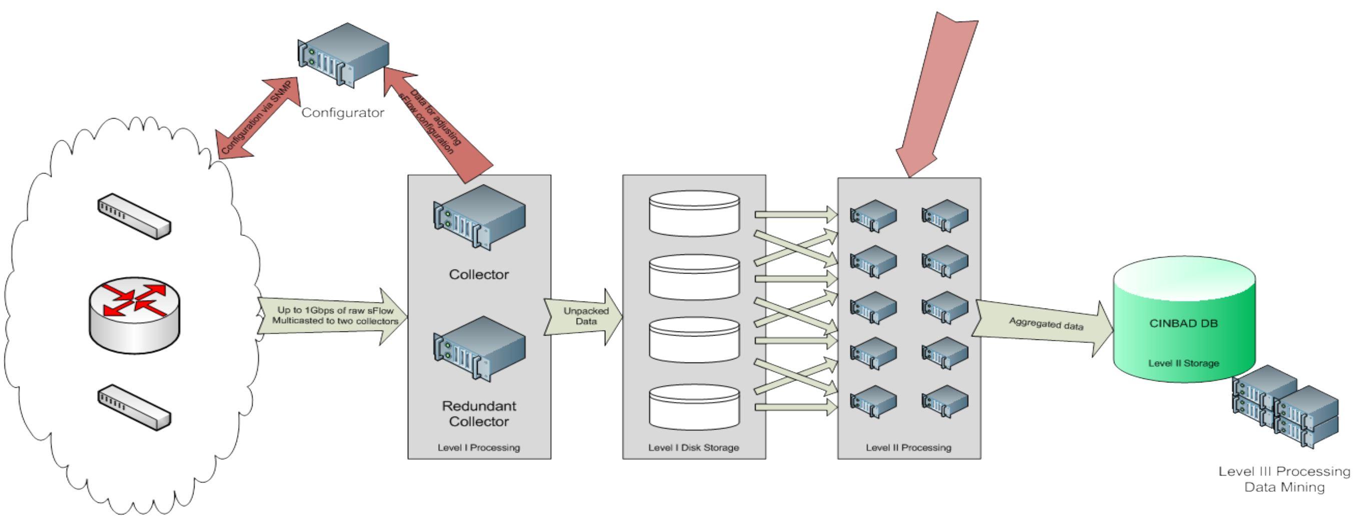
Ryszard Erazm Jurga - CERN
Milosz Marian Hulboj - CERN

Outline

- Update on
  - CINBAD data collection
  - Anomaly detection
- CINBAD enhancements for CERN Network Monitoring
- Collaboration, Publications and Presentations

CINBAD data collection

- Current collection based on the traffic from ~1000 switches
  - ~100GB data storage per day
  - ~6000 sampled packets per second
  - ~3500 snmp counter samples per second

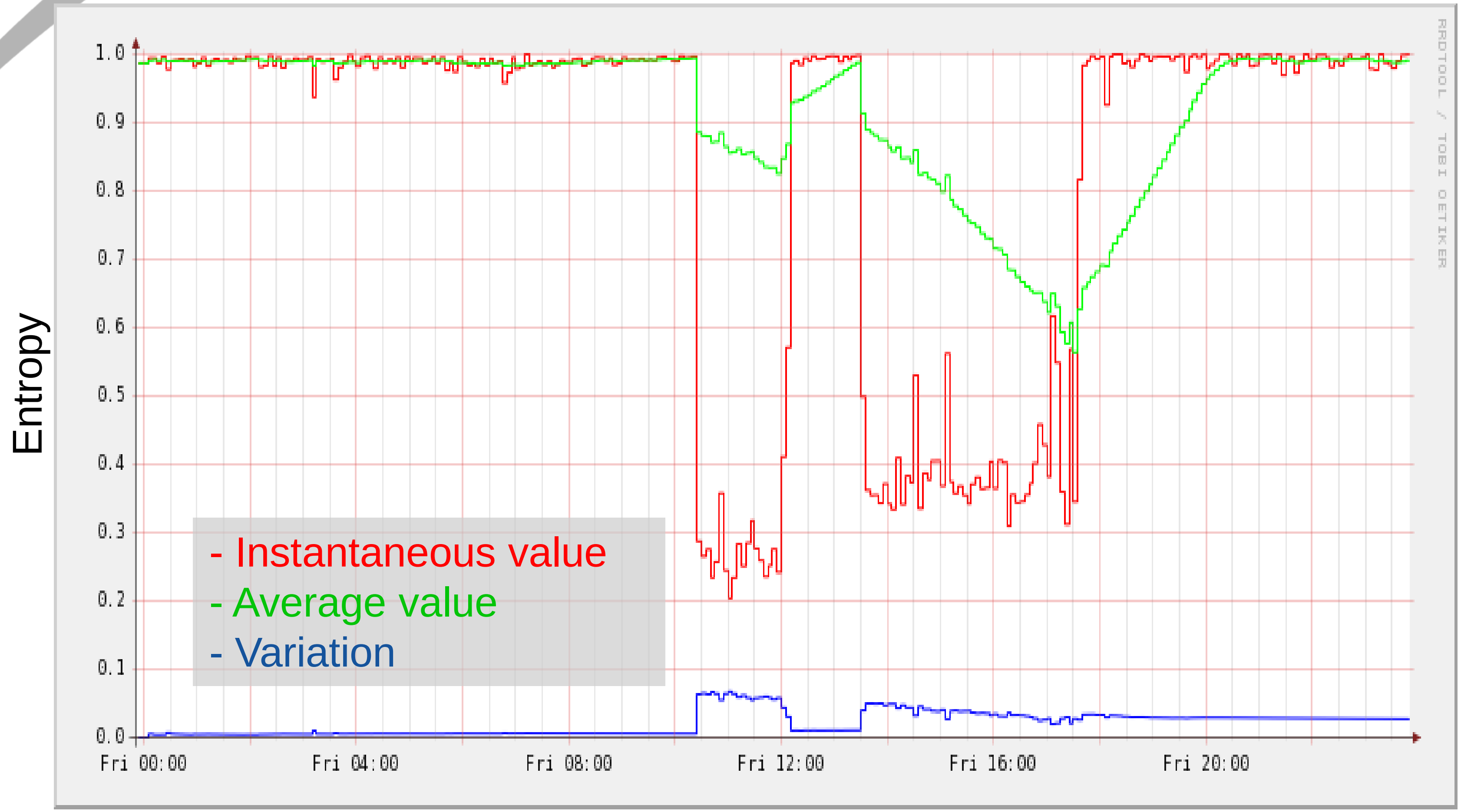


Anomaly detection update

- Both statistical analysis and pattern matching techniques in use
  - internal and external traffic analysed
- Anomalies found
  - DDoS client
  - Conficker infections
  - Spammers
  - non-legitimate CERN-wide network scans and external scans (e.g. ~40M addresses contacted on port 445 by one CERN host over 2 days)
  - viruses and keyloggers

Conficker infection

Entropy of the number of distinct destination addresses per tcp dst port



CINBAD for Network Monitoring

- Brainstorm sessions organized to identify areas for potential improvements
  - host activity and connectivity record
    - where the host is connected to?
    - who the host communicated with?
    - what type of traffic does the host send?
    - …
  - link utilization visualization
    - port exhibiting utilization above x%
  - Network statistics and trends
    - e.g. #flows, #active ports, traffic volume
    - average number of hosts per switch port
  - need for post mortem analysis facilities

Data Visualisation I

- CINBAD gathers different statistical information about the network
- Much of the data has **hierarchical nature**
- Need for generic visualisation tool

- Defined a summer student project:
  - Examine the available libraries
  - Adapt to CINBAD needs
  - Provide sample applications

Data Visualization II

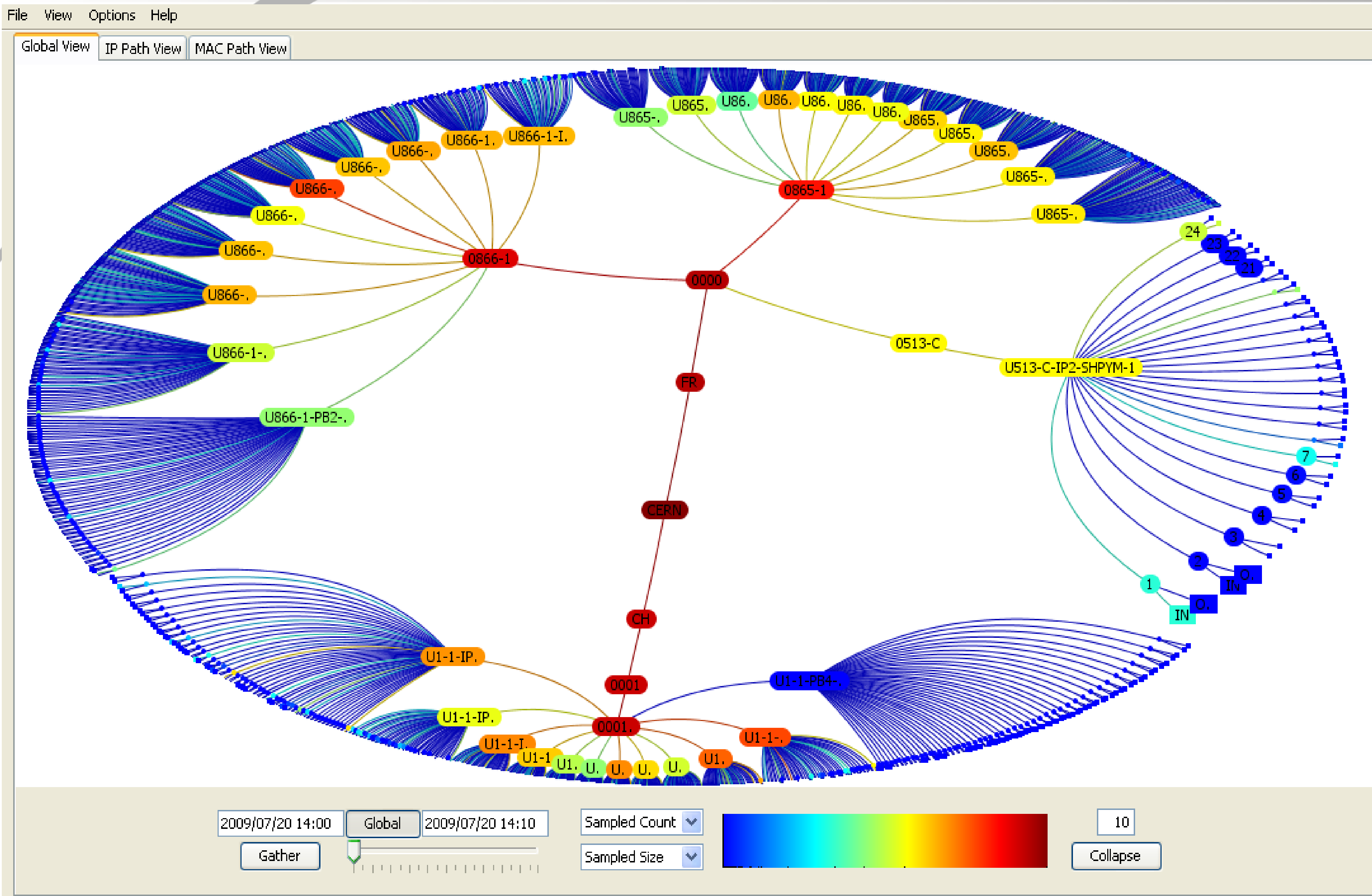
- Investigation of library suitable for visualization of hierarchical data

- Successful adaptation of Treeviz library by our summer student – Vlad Petre

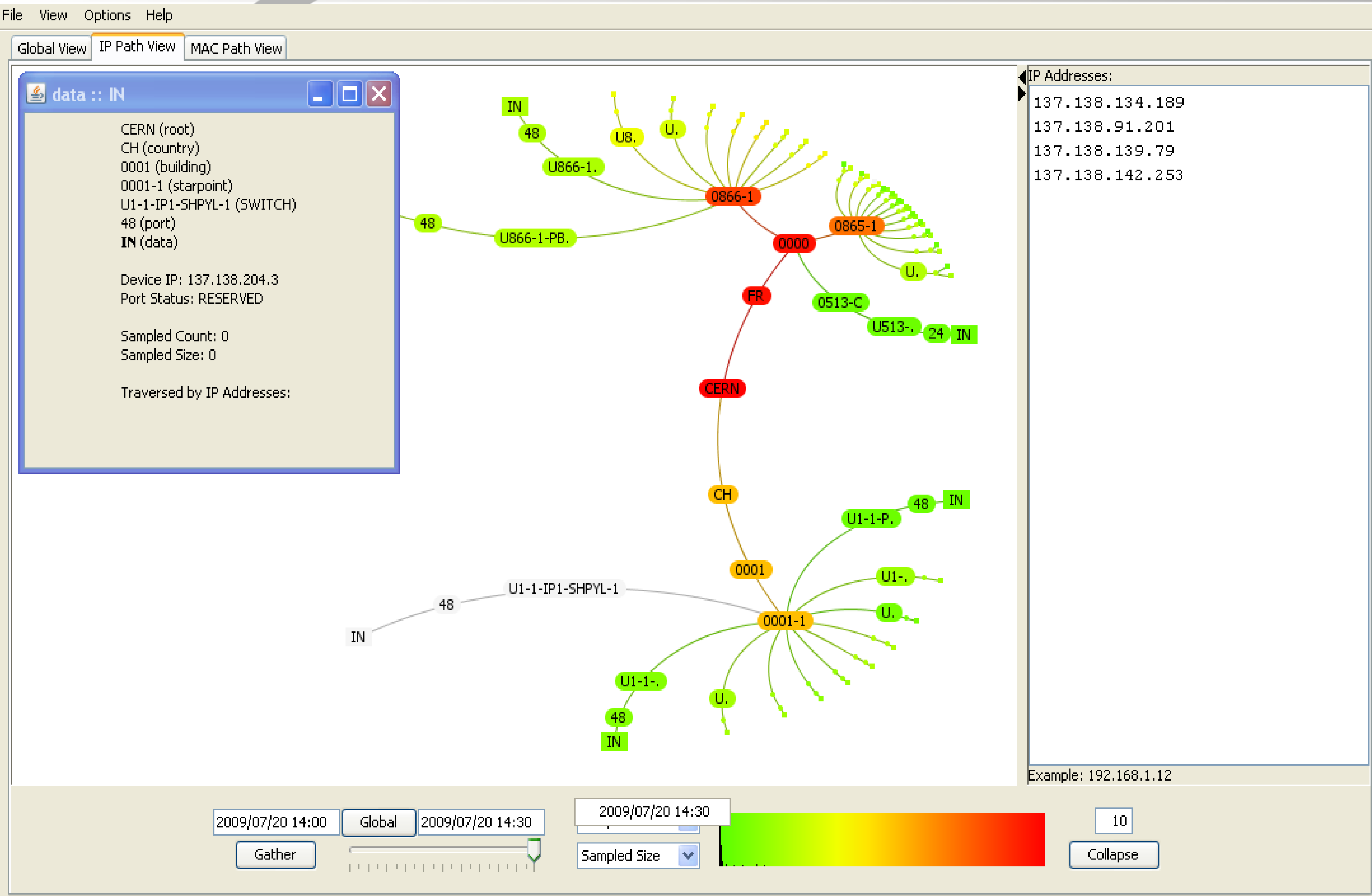
- Detailed project report delivered

- Useful applications:
  - Visualization of the sFlow collection status
  - Tracing the ports on which given Ethernet/IP address was seen

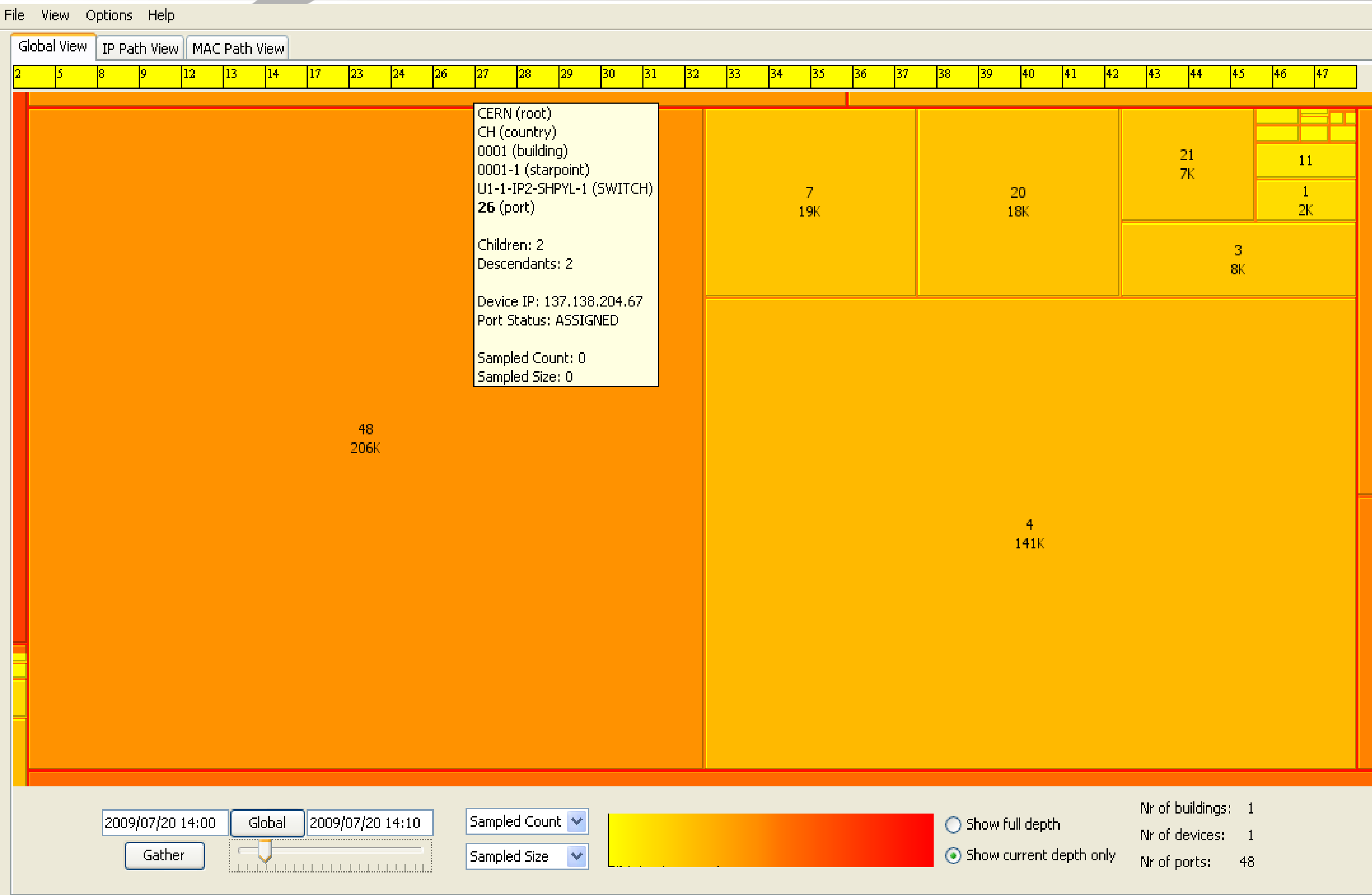
Hierarchical sFlow collection status



Tracing Ethernet/IP Addresses



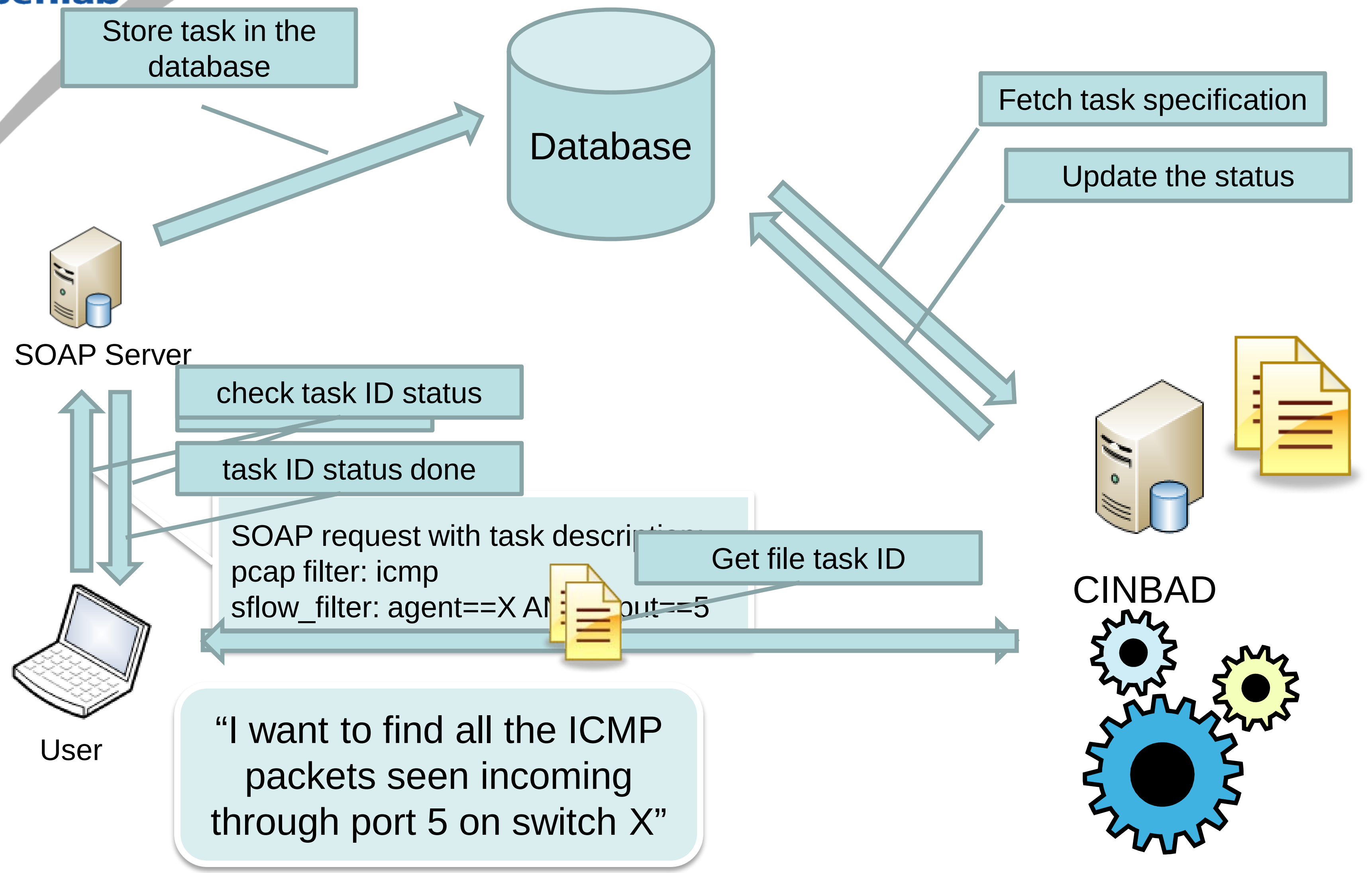
sFlow Collection Status



CINBAD CERN-wide tcpdump

- Sampled traffic stored in one place
- Data is confidential
- Specialised tool provides information
  - about the sampled packets
  - where and when the traffic has been seen
- Particularly useful in detecting packets '*that should not be there*' (policy violation)
- Examples of applications:
  - Find specific kind of traffic (i.e. rogue DHCP)
  - Filter all the traffic between the set of machines (e.g. PLCs and the rest of the network)

Access to CINBAD tcpdump



Collaboration with UNIRIO

- Request for data traces from CERN network made by UNIRIO (Universidade Federal do Estado do Rio de Janeiro) researchers
  - investigation of the entropy metric usage for anomaly detection
  - came via Procurve Networking

- Anonymized data traces have been provided
  - a special tool developed by CINBAD to anonymize data
  - worm infection included and labeled

- The UNIRIO analysis was not capable of detecting the worm
  - potential reason: transition from sflow to netflow flows

Publications and Presentations

- HEPiX Presentation, May 26th
- Post-C5 presentation, June 12th
- CNL July-September 2009
  - CINBAD keeps an eye on the CERN network
  - front page article
- Recent Advances in Intrusion Detection (RAID) Conference, September 23th
  - poster
- Contributing to HP Tech Con '10
  - the focus is on technical innovation, HP internal conference

CERN COMPUTER NEWSLETTER

Volume 44, issue 3 July–September 2009

Contents

Editor Natalie Pocock, CERN IT Department, 1211 Geneva 23, Switzerland. E-mail cnl.editor@cern.ch. Fax +41 (22) 766 8500. Web cerncourier.com/articles/cnl.

Produced for CERN by IOP Publishing Dirac House, Temple Back, Bristol BS1 6BE, UK. Tel +44 (0)117 929 7481. E-mail jo.nicholas@iop.org. Fax +44 (0)117 930 0733. Web iop.org.

Published by CERN IT Department

IOP Publishing

CINBAD keeps an eye on the CERN network

The CINBAD (CERN Investigation of Network Behaviour and Anomaly Detection) project was launched in 2007 as a collaboration between CERN openlab, IT-CS and HP ProCurve Networking. The project's aim is to understand the behaviour of large computer networks in the context of high-performance computing and campus installations such as those at CERN. The goals are to detect traffic anomalies in such systems, perform trend analysis, automatically take counter measures and provide post-mortem analysis facilities.

CERN's network

CERN's campus network has more than 50000 active user devices interconnected by 10000 km of cables and fibres, with more than 2500 switches and routers. The potential 4.8 Tbps throughput within the network core and 140 Gbps connectivity to external networks offers countless possibilities to different network applications. The bandwidth of modern networks is growing much faster than the performance of the latest processors. This fact combined with the CERN specific configuration and topology makes network behaviour analysis a very challenging and daunting task.

CINBAD in a nutshell

The CINBAD project addresses many aspects associated with the CERN network. First, it provides facilities for a better understanding and improved maintenance of the CERN network infrastructure. This includes analysing various network statistics and trends, traffic flows and protocol distributions. Other factors that might have an impact on the current network status or influence its evolution are also studied, such as connectivity, bottleneck and performance issues.

When we have learnt and understood the network behaviour, CINBAD can help to identify various abnormalities and determine their causes. Because there are many factors that can be used to describe the network status, anomaly definition is also very domain specific and includes network infrastructure misuse, violation of a local network security policy and device misconfiguration. In addition, the expected network behaviour never remains static because it can vary with the time of day, the number of users connected and network services deployed. As a consequence, anomalies are not easy to detect.

Network sniffing

To acquire knowledge about the network status and behaviour, CINBAD collects and analyses data from numerous sources. Alarms from different network monitoring systems, logs from network services like Domain Name System (DNS), Dynamic Host Configuration Protocol (DHCP), user feedback, etc – all of these constitute a solid base of information. A naive approach might be to look at all of the packets flying over the CERN network. However, if we did this we would need to analyse even more data than the LHC could generate. The LHC data are only a subset of the total data crossing via these links.

CINBAD overcomes this issue by applying statistical analysis and using sFlow, a technology for monitoring high-speed switched networks that provides randomly sampled packets from the network traffic. The information that we collect is based on the traffic from around 1000 switches and routers and gives a representative sample of the CERN network traffic with more than 3 Terabytes of data per month. The multistage collection system was designed and implemented in consultation with experts from the LHC experiments and Oracle, to benefit from their data-analysis and storage experience. The system has now been up and running for more than a year (figure 1).

Network operation enhancements

The field of network monitoring and planning can greatly benefit from the CINBAD activities. We provide tools and data that simplify the operation and problem-diagnosing process. In addition, our statistics help in understanding the network evolution and design.

http://cern.ch/openlab-cinbad

RAID 2009
12th International Symposium
On Recent Advances In Intrusion Detection
Saint-Malo, Brittany, France | September 23-25, 2009

CERN Investigation of Network Behaviour and Anomaly Detection

Milosz Marian Hulboj and Ryszard Erazm Jurga
{mhulboj,rjurga}@cern.ch
CERN – HP Procurve openlab project

The CINBAD (CERN Investigation of Network Behaviour and Anomaly Detection) project was launched in 2007 in collaboration with ProCurve Networking by HP. The project mission is to understand the behaviour of large computer networks in the context of high performance computing and large campus installations such as at CERN, whose network today counts roughly 70,000 Gigabit user ports. The goals of the project are to be able to detect traffic anomalies in such systems, perform trend analysis, automatically take counter measures and provide post-mortem analysis facilities.

With the modern high-speed networks it is impossible to monitor all the packets traversing the links. sFlow is the industry standard for monitoring high-speed switched networks overcomes this issue by providing randomly sampled packets (first 128 bytes) from the network traffic.

sFlow is scalable and can monitor links of all speeds without impacting the performance of the network devices. It is a low cost solution that is supported by a wide range of vendors.

At CERN we collect the sFlow data from more than 1000 switches and routers. Per month we gather more than 3TB of data.

Initial bytes of data obtained by sFlow provide a centralised network-wide view of the network activity. To give an illustration, one could compare it to network-wide tcpdump that collects randomly sampled packet headers from the whole network providing additional metadata at the same time.

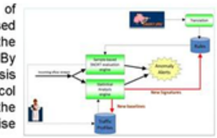
Credit to IOP Publishing

Apart from providing a useful debugging information for the network engineers, collected packets are crucial for the analysis conducted by the CINBAD team.

We have been investigating various data analysis approaches that could be categorised mainly into the two domains: statistical and signature based analysis. The former depends on detecting deviations from normal network behaviour while the latter uses existing problem signatures and matches them against the current state of the network. The signature based approach has numerous practical applications with SNORT being a prominent example.

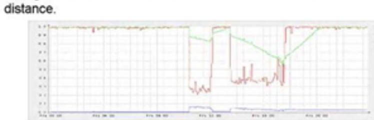
The CINBAD team has successfully ported SNORT and adapted various rules in order to work with sampled data. It seems to perform well, and provides a low false positive rate. However, the system is blind and can yield false negatives in case of unknown anomalies.

Fully automatic detection of novel worms and unsupervised signature generation is the unattainable *Holy Grail*. By combining the statistical analysis of the data from the protocol headers and by analysing the payload we attempt to minimise the necessary supervision.



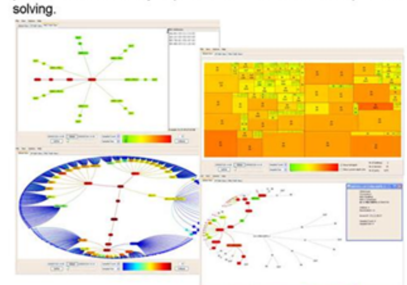
The behaviour of each host is described by its profile. A set of all host profiles is considered as a network profile. This is the most natural way to represent the network, since each host partially influences the network behaviour and we can directly point out dominant hosts in case of an anomaly.

Each profile consists of features originating from packet headers (e.g. number of distinct TCP ports contacted in Δt, ratio of egress/ingress traffic of a given type) or packet payload (e.g. entropy of the payload, *n*-gram distribution, hashes of the partitioned payload). For each profile we set up a baseline using historical data and then compare those baselines against latest ones. Different distance metrics are employed, most notably divergence, standardised Euclidean distance and Mahalanobis distance.



The Conficker worm infection identified by the CINBAD team and visible as a double drop in one of the metrics used

CINBAD team had also developed (with the help of student Vlad Petre) several visualisation tools for CERN network engineers to help them with daily work. Network-wide visibility proves to be extremely important for maintenance and problem solving.



References:

[1] Jurga R., Hulboj M., *Technical Report Packet Sampling for Network Monitoring*, CERN, 2008.
[2] Kim H., Karp B., *Autograph: Toward Automated, Distributed Worm Signature Detection*, USENIX, 2004.
[3] Wang K., Stolfo J., *Anomalous Payload-Based Worm Detection and Signature Generation*, RAID 2005.

ProCurve Networking by HP

Conclusions and plans

- Statistical analysis with pattern matching provides encouraging results in anomaly detection
  - The technical report about anomaly detection techniques will be sent by the end of this week

- CINBAD can provide useful enhancements to CERN Network Monitoring
  - Prototype of CERN tcpdump will be available for IT-CS in October
  - Identification of needs and development of other tools for IT-CS Network Engineers in the next weeks