CERN COMPUTER NEWSLETTER

Volume 44, issue 3 July–September 2009

Contents

Editorial
CINBAD keeps an eye on the CERN network 1
ETICS 2 offers guidance to software professionals 3
Announcements and news
CERN welcomes 13 Intel ISEF pre-college winners 4
Computer team advises reviewing your security now and frequently 5
EGEE-III project is on track for EGI transition 5
Grid news
Scientists demonstrate the role of CMS in computing Grid 6
Technical brief
Indico's new face goes live 7
CERN updates Wi-Fi network 9
Conference and event reports
Prague hosts CHEP conference 10
Workshop identifies steps to reap benefits from multicore and virtualization technologies 11
HEPiX event arrives in Sweden 12
Calendar 12

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.

Advisory board Frédéric Hemmer (head of IT Department), Alberto Pace (group leader, Data Management), Christine Sutton (*CERN Courier* editor), Tim Smith (group leader, User and Document Services).

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

©2009 CERN

The contents of this newsletter do not necessarily represent the views of CERN management.

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 50 000 active user devices interconnected by 10 000 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.

A very basic piece of information that is of interest for network operations is knowledge about the host's activity. CINBAD is able to provide detailed statistics about the traffic sent and received by a given host, it facilitates inference about the nature of the traffic on a given outlet/port and can thus identify the connected machine. This information could also be used to diagnose routing problems by looking at all of the packets outbound or inbound to a particular host.

CINBAD is also able to provide information about the traffic at CERN. The sampled data collected by the project are sufficient to obtain the switching/routing/transport protocol information as well as gaining information about the application data. This provides valuable input for an understanding of the current network behaviour. Here the CINBAD team uses descriptive statistics. The potential set of metrics that we can provide to characterize the traffic at CERN is very extensive and specific needs are currently being discussed. For example, we can enumerate protocol-type distributions, packet size distributions, etc. Depending on the requirements, these statistics can be tailored even further.

Top n-list is another form of network summary that might be of interest. Such lists would allow the identification of the most popular application servers, either inside or outside CERN. Although this information might be available on each individual CERN server, CINBAD provides the possibility to collect these statistics for all servers of a given type, whether or not they are centrally managed by the IT Department. This information may be of value to both network engineers and application-server administrators.

These statistics can also be useful for network design and provisioning. The CINBAD project can provide valuable information about the nature of the traffic on the links. These statistics can also be used to detect the trunks with potential bottlenecks. This information can be compared with the service-level agreements that specify the conditions for link usage, enabling appropriate corrective actions to be taken.

With all of these improvements, CINBAD offers a comprehensive system to facilitate day-to-day operations, diagnose network problems and extend our understanding of network evolution and design. The CINBAD team is currently working in close collaboration with IT-CS on a visualization model of this information that is suitable for network operation and troubleshooting.

Security enhancements

Security is another area that benefits from the CINBAD project. The only safe computer is a dead computer, or at least one disconnected from the network (if no-one can get to it, no-one can harm it). Nowadays, we cannot avoid communicating with others and therefore we expose our machine to outside threats. Although CERN centrally managed desktops have up-to-date anti-virus software and firewalls, this does not guarantee that our machines and data are shielded from attacks. These tools are usually designed to detect known patterns (signatures) and there are also other machines (unmanaged desktops, PDAs, etc) connected to the CERN network that might be less protected.

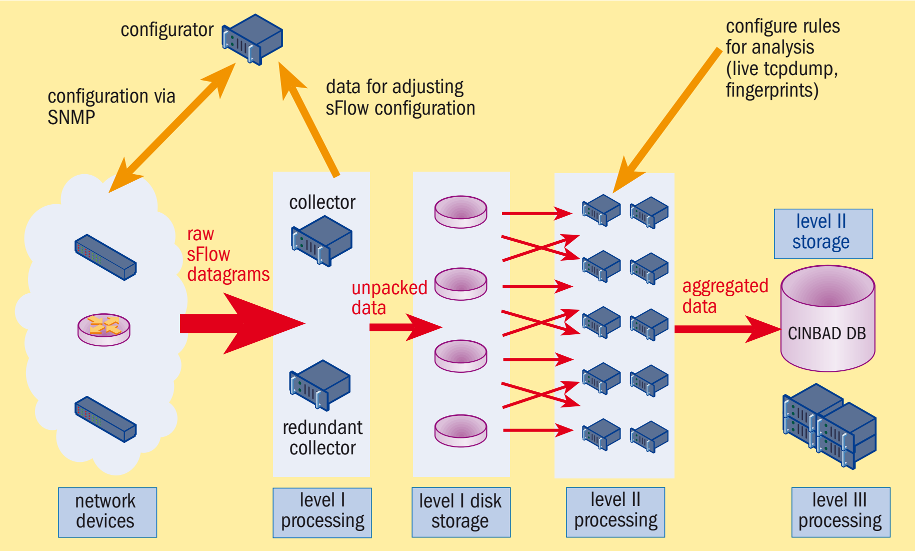

Fig. 1. The CINBAD sFlow data collector receives and processes the CERN network traffic.

Currently, detailed analysis is only performed at critical points on the network (firewall and gates between network domains). The CINBAD team has been investigating various data-analysis approaches that could overcome this limitation. These studies can be categorized into two main 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, for example SNORT (an open-source intrusion-detection system). The CINBAD team has successfully ported SNORT and adapted various rules 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 cases of unknown anomalies.

This problem can be addressed by the statistical approach. Expected network activity can be established by specifying the allowed patterns in certain parts of the network. While this method works well for a DNS or web server that can only be contacted on a given protocol port number, for more general purposes this approach would not scale.

A second approach is to build various network profiles by learning from the past. The selection of robust metrics that are resistant to data randomness plays an important role in characterizing the expected network behaviour. Once these normal profiles are well established, the statistical approach can detect new and unknown anomalies.

The CINBAD project combines the statistical approach with the signature-based analysis to benefit from the synergy of the two techniques. While the latter provides the detection system with a fast and reliable detection rate, the former is used to detect the unknown anomalies and to produce new signatures. The CINBAD team constantly monitors both the campus and internet traffic using this method. This has already led to the identification of various anomalies, e.g. DNS abuse, p2p applications, rogue DHCP servers, worms, trojans, unauthorized wireless base stations, etc. Some of these findings have resulted in refinements to current security policies.

The future

The CINBAD project offers many opportunities to improve CERN's network operation, and it also provides a unique opportunity for the CERN Computer Security Team to identify (and protect against) incidents that might not be seen otherwise. It also enables other groups concerned with varying network applications, such as web services and mail servers, to understand their behaviour.

Useful links

The CINBAD project: http://cern.ch/openlab-cinbad
CERN openlab: http://cern.ch/openlab

Milosz Hulbój and Ryszard Jurga, IT-CS (CERN openlab)