MASTS : Measurements at All Scales in Time and Space

Jordan Augé, Richard Clayton, Andrew W. Moore

UNIVERSITY OF CAMBRIDGE
800 YEARS
1209~2009

Introduction

The aim of the MASTS project is to build continuous-capture streaming monitoring systems and operating them inside the JANET network. MASTS provides anonymized datasets of packets to the research community in real time. JANET is the National Research and Education Network for UK. It of ers Internet connectivity to universities, institutions of education and research organizations and thus transports various types of traffi from ISP and GRID/ eScience, for example data from the Large Hadron Collider located at CERN. While such infrastructure might appear easy to build, it provides a considerable challenge in legal and engineering terms.

Motivation

Real network data is scarse but important in network research. While some high-speed *snapshot* monitors are available (nProbe, GridProbe, etc.), providing streaming capture data in real-time opens new opportunities for studies, like the design of performance and diagnostic systems, low-error Intrusion Detection and Information Assurance systems.

The engineering challenges

The system had to fit in a colocated environment, and integrate inside an operational network with minimal disruption, power or space needs.

The software is written with **robustness** in mind, to capture and anonymize data continuously at 10Gb/s.

Sustaining high data throughput, especially in the WAN environment between the capture and processing point, was made possible via a SAN (Storage Area Network) operated with GFS (Global File System), and adequate buffering to compensate for throughput variations.

Storage of such **large datasets** is done by keeping data for differen time periods, according to their detail level.

An industry standard algorithm allows for **prefix-preserving anonymization of IP addresses**. This maintains the IP structure, allowing for studies of routing and identifying groups of end-systems while preserving their anonymity.

The legal challenges

Reconciling the needs of the **different participants**: network users, network operators and network researchers.

Designing a l**egal framework** to allow for the lawful interception and dissemination of packet headers...

...while providing a sufficien level of **anonymization to protect users yet be useful for research purposes.**

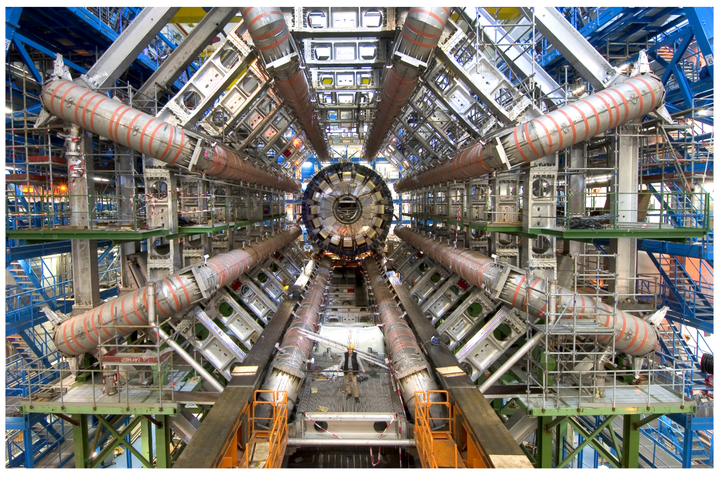

Fig.2 - ATLAS detector at LHC (Large Hadron Collider) in CERN. One of the sources of data stored in the UK and monitored by the MASTS project.

The probes

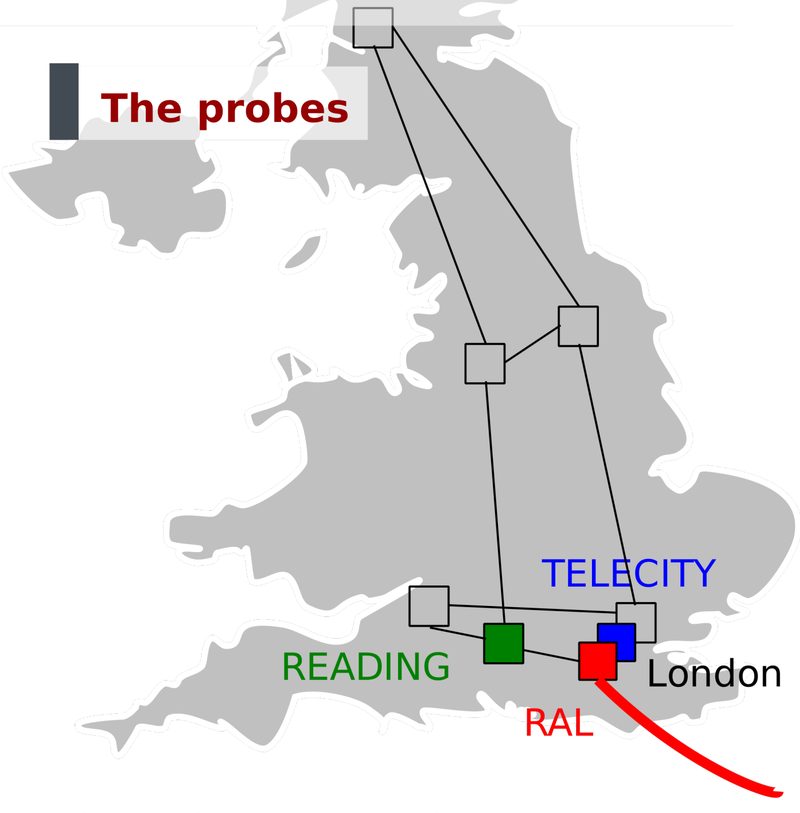

Fig.1 - The MASTS probes are deployed in those 3 core points of presence of the JANET network.

- the TELECITY probe captures the traffi from a lightpath inside the core network;
- in READING, two probes monitor an internet interconnection point;
- the RAL endpoint is connected to the LHC/CERN infrastructure and two probes receive full-duplex data *(see Figure 2).*

The data

The anonymized datasets of packet headers at **layers 2,3 and 4** are indexed along with **metadata** in a database, and made available through the project website to users after the signature of an Acceptable User Policy.

Streaming capture architecture

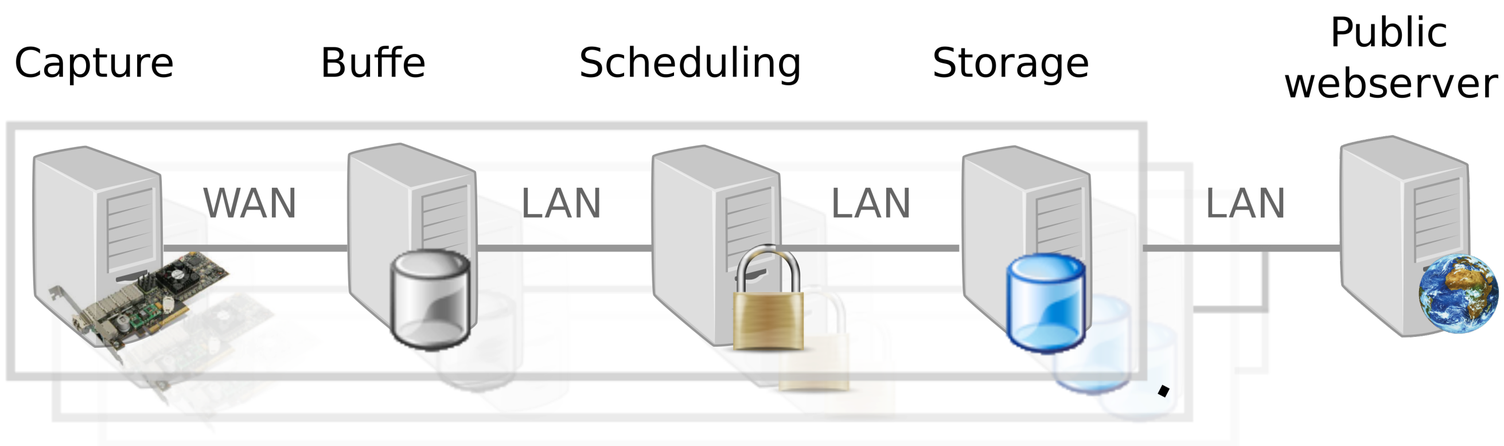

Fig.3 - Once collected, the data is anonymized and transported up to a permanent store, where it is indexed and made available in near-real time to research users.

References

- R. G. Clegg *et al.* “Challenges in the capture and dissemination of measurements from high-speed networks”, *IET Communications, 2008*
- *other publications about trace compression and analysis are available online*

Project: http://www.mastsproject.org

Data: http://data.mastsproject.org

UNIVERSITY OF CAMBRIDGE · UCL · Loughborough University · ja.net · endace power to see all · EPSRC