Marek Dąbrowski, Jarosław Śliwiński, Wojciech Burakowski
Institute of Telecommunications
Warsaw University of Technology, Warsaw

The MOME Measurement Database

In this paper we present the measurement database developed by IST project MOME (Monitoring and Measurement Cluster). The MOME project is a Coordination Action (CA), with a goal to harmonize the activities of European projects working in the area of IP networks monitoring and measurements. One of the tasks of MOME is to maintain the repository of information about publicly available measurement data, like e.g. packet and flow traces, or routing data. The access to raw measurement data is crucial for many different areas of research related with Internet traffic. The MOME repository aims at assisting the research community in finding and retrieving useful and representative data. In this paper we discuss the MOME database model and outline additional capabilities related with performing automated analysis of stored measurement data. Moreover, we show exemplary results of analysis performed on the traces from publicly available trace archive.

1. Introduction

It is commonly believed, that the Internet traffic is still not well understood and difficult to be described by mathematical formulas. Developing and validating new realistic traffic models (see e.g. [3][4]) relies on access to repositories of real measurement data. Notice, that a number of research projects aims at collecting measurement data (see e.g. [5][6][7][8][9]), but typically each one does it on its own way and using different measurement tools. Due to the lack of standard formats for data storage and annotation, finding appropriate data among various uncoordinated repositories is often quite difficult.

The Coordination Action IST-MOME (Monitoring and Measurement Cluster) [1] has developed a meta-database, which collects information about measurement data available from different repositories. The format of stored information (i.e. "meta-data model") allows for describing the measurement scenario and environment by a carefully selected set of attributes [2]. Notice, that the meta-database approach has been previously studied by several projects dealing with measurement data. For example, the SIMR architecture [10] and CADIA Trends project [11] aim at collecting and annotating various measurement data. The MOME database not only stores the information, but can also perform automated analysis of submitted measurement data. The MOME Data Analysis Platform allows for executing selected analysis tools. The access to the results is available for all visitors browsing the MOME database. Together with meta-data describing the assumed measurement scenario, the analysed results may help to assess, if a particular raw data set is appropriate for given research targets.

The paper is structured as follows. In section 2 we introduce the MOME meta-data model. In section 3 we shortly describe the MOME Data Analysis Platform. In section 4 we outline the deployment and current status of MOME database. The exemplary case study, which demonstrates the capability of MOME system for performing selected analysis tasks, is presented in section 5. Finally, section 6 concludes the paper.

2. The MOME meta-data model

In this section we describe the meta-data model, which was used as a base for implementation of the MOME meta-repository. The meta-data model consists of a set of attributes, which annotate the most important features of various measurement data. These attributes describe in detail the measurement scenario, like e.g. the type of measurement, type and format of stored data, type of capture platform. Additionally, the meta-data model comprises some statistical parameters, which represent the internal characteristics of measured traffic. Currently five different types of measurement data are considered, as described below: Packet traces, flow traces, QoS results, routing data and HTTP traces. In addition the MOME database supports a generic type for web repositories having mainly a pointer to a base URL.

2.1 Packet traces

The packet traces consist of collections of time-stamped records of packets, captured in certain measurement point. Obviously, the availability of representative packet traces is crucial for developing and validating packet-level traffic models.

Notice, that each trace is captured at a certain point of a network, only within a finite time interval. The trace can be regarded as representative, if its relevant features sufficiently reflect the general characteristics, and can be confirmed by a trace collected at a different time, or in different network. For example, different traffic models are proposed for the access and core networks, which can differ essentially from the point of view of level of traffic aggregation. For evaluating the traffic model, developed for core or access network, the used packet traces must be representative for the assumed type of network. Additionally, the duration of the trace should be sufficiently long to allow evaluating the stationary behaviour of captured traffic (or, to reveal its non-stationary behaviour).

To allow assessing if the trace is representative and appropriate, the information stored in the MOME meta-data attributes should include such information, as: type of measured network, measurement time, location of capture device, speed of monitored link, capture method, filtering and anonymisation rules, capture platform, and the storage format of collected data. This information gives us detailed description of the assumed measurement scenario. The MOME meta-data model attributes defined for packet traces are presented in Table 1.

Table 1. The packet traces attributes

Detail attributes
- Type of network, where the trace was captured (e.g. LAN or WAN),
- Location of the capture device,
- Type of captured traffic (e.g. if the trace was taken from operational network or from controlled testbed),
- Type and speed of the monitored link,
- Capture method (e.g. if the trace was recorded directly by the router, or by external capture device),
- Filter rules (i.e. if the trace contains all observed packets, or only a subset, defined according to certain rules),
- Number of captured packets,
- Contents of the record (i.e. if the trace contains entire TCP/IP headers with timestamps, or only selected header fields),
- Trace anonymisation rules (i.e. if the IP addresses were removed from the trace or scrambled to protect privacy of users),
- Capture platform (e.g. if it was Linux router or DAG card),
- Data format (e.g. [13]: *Libpcap*, *DAG*, *tcpdump*, *PSAMP*, *sFlow*),
- Additional information

Table 2. The flow traces attributes

Detail attributes
- Type of network, where the trace was captured,
- Location of the capture device,
- Type of captured traffic (e.g. if the trace was taken from operational network or from controlled testbed),
- Type and speed of the monitored link,
- Capture scenario (e.g. if the trace was recorded directly by the router, or by external capture device attached to the link),
- Filter rules (i.e. if the trace contains all observed flows or only a subset, defined according to certain rules),
- Number of captured flows,
- Trace anonymisation rules (i.e. if the IP addresses were removed from the trace or scrambled to protect privacy of users),
- Capture platform (e.g. if it was Linux router or DAG card),
- Data format (e.g. [13]: *IPFIX*, *NetFlow*, *IPDR*)
- Additional information.

2.2 Flow traces

The flow traces consist of a collection of records of distinct flows (usually identified by the matching values of proper IP header fields), observed in certain measurement point. The flow record contains the timestamps of flow arrival and departure, as well as the number of packets and bytes transmitted within its lifetime. The flow traces are useful in developing and validating the traffic models on the flow (or connection) level. Similarly to the case of packet traces, the capture location and duration should be considered for assessing the usefulness of the trace. Therefore, the following meta-data attributes have been defined: type of measured network, measurement time, location of capture device, speed of the monitored link, capture method, filtering and anonymisation rules, type of capture platform, and the storage format. The MOME meta-data model attributes defined for packet traces are presented in Table 2.

2.3 QoS results

In the case of QoS measurements, the raw data consists of a set, or time-series, of measured values of QoS metrics, like e.g.: one way delay, delay variation, packet loss ratio, etc. The singleton values correspond to the measurement taken on a single packet (probe packet in the case of active measurement, or data packet in the case of passive measurements).

The meta-data attributes should give precise information on the network, where the measurement was performed, and the time, when it was performed. Therefore, the defined meta-data model includes such information, like: the name and type of network, location of measurement equipment, method for time synchronisation between the transmitter and receiver, calculated metrics, and the storage format of the measurement results. The MOME meta-data model attributes defined for packet traces are presented in Table 3.

Table 3. The QoS results attributes

Detail attributes
- Type of network, where the measurement was performed, - Measurement method (active or passive), - Measured metrics (e.g. OWD, IPDV, RTT, packet losses…), - Location of the measurement points, - Measurement platform, - Time synchronization method (e.g. NTP or GPS), - Number of measured singleton values, - Data format, - Additional information.

Table 4. The routing data attributes

Detail attributes
- Routing protocol (e.g. OSPF or BGP), - Type of collected data (i.e. snapshots of routing tables, or collections of routing update messages), - Location of collector device, - Data format (i.e. *Zebra* or *Tcpdump*),

Table 5. The HTTP traces attributes

Detail attributes
- Location of the capture device, - Filter rules, - Number of captured entries, - Trace anonymisation, - Capture platform, - Data format, - Additional information.

2.4 Routing data

The routing data consists of the snapshots of routing tables, or collections of routing update messages captured within particular measurement interval. The routing data may be useful for investigating the changes of logical inter-domain topologies, observing trends in number of advertised prefixes, or detecting abnormal behaviours of the inter-domain routing. The MOME meta-data model attributes defined for packet traces are presented in Table 4.

2.5 HTTP traces

The HTTP (Hyper Text Transfer Protocol) is the application layer protocol, designed for the web application. The HTTP trace contains records of observed user HTTP requests and server HTTP responses typically collected by log-files in web- or proxy-servers. The information stored in such traces corresponds to the application (or session) layer traffic model. The MOME meta-data model attributes defined for packet traces are presented in Table 5.

3 Automated analysis of measurement data

The MOME Data Analysis Workstation allows the users to request selected analysis tasks on the measurement data, which is annotated in the MOME database. The raw data file is retrieved from its original location and temporarily stored it on a local disk of MOME data analysis workstation. After performing the analysis, the results are inserted into appropriate fields in the MOME Database.

The MOME system assumes that external, freely available data analysis tools can be used for performing actual operations on the measurement data. Each tool must be adapted to the MOME Workstation by implementing appropriate filters, which are able to read and parse the analysis results. Figure 2 illustrates the internal structure of software modules composing the MOME Data Analysis Workstation, and the step-by-step process of executing the data analysis task. The process is initiated by the Analysis Request Manager (ARM), which is a part of MOME GUI. The Autonomous Process Manager (PM) is periodically executed on the MOME Database server and searches the database for new pending analysis tasks. If PM finds out that new analysis task has been requested, it starts the Download Manager (DM) (step 2). The DM downloads the raw measurement data file from specified location and saves it on the local storage (step 3).

Then, the PM starts the Analysis Manager (AM) (step 4), which executes the actual external analysis tool (step 5). Remark, that the tool has to be first installed with appropriate privileges on the MOME server. After finishing the actual analysis work, the analysis tool saves the results into the output file (step 6). The AM parses this file and inserts appropriate values into the MOME database (step 7).

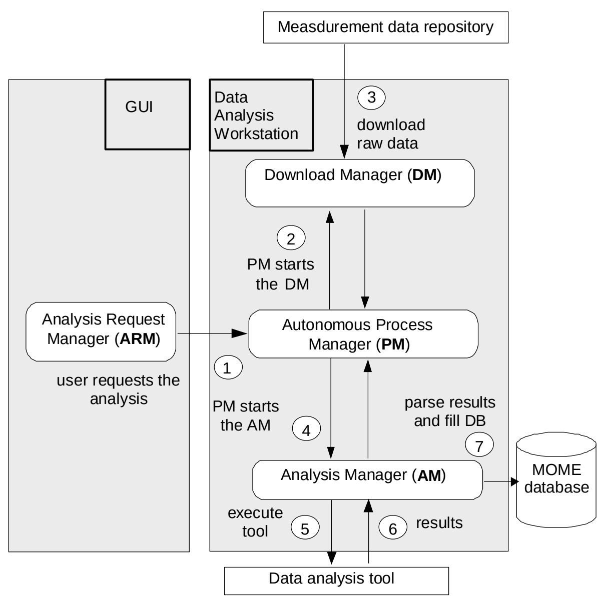


Figure 1. MOME Data Analysis Workstation

4. Deployment of MOME database

The measurement database, as well as the MOME Data Analysis Workstation, were implemented and deployed on a server located in the premises of one of the project partners. The database is running and it is publicly accessible via the web-page [1] for submitting data and performing data analysis.

Currently, the main “source” of data for MOME measurement database is the packet capture point located in the FH Salzburg in Austria. The traffic outgoing from the local network of this educational institution is captured 23 hours a day, in 1-hour long intervals. At midnight, the data captured during the day are transferred to the MOME database, together with automatically generated values of the meta-data attributes. At the time of writing this paper, more then 2400 1-hour long traces were available in the MOME database, representing 60GB of captured traffic.

5. Example of MOME data analysis for packet traces

Several data analysis tools were integrated into the MOME Data Analysis Workstation. Currently they allow for performing basic analysis tasks for packet-level traces that are stored in the formats of libpcap [12] and DAG [14]. However, notice that new analysis tools, aimed at supporting different analysis tasks and data formats, can be quite easily added. Below, we present the exemplary results of analysing one of the traces annotated in the MOME database, that is the publicly available trace from the MAWI archive [9].

The attributes of the MOME meta-data model, as specified in section 2.1, are presented in Figure 2. One can observe, that the quick overview of available attributes allows for assessing the basic characteristics of traffic captured and stored in a particular trace.

Dataset name	MAWI Samplepoint A, 31.01.2005
Data type	PacketTrace
File size	448.2MB
File compression	gz
Start time	2005-01-31 14:00:01
End time	2005-01-31 14:15:01
Duration	15min
Description	Wide-area traffic captured on USA-Japan link. Trace is stored by MAWI Working Group Traffic Archive
Dataset location	URL: ftp://tracer.csl.sony.co.jp/pub/maw...lepoint-B/2005/200501311400.dump.gz
Tool	Tcpdump
Submitted by	mdabrowski @ 2005-02-01 18:51:49
Last Update	2005-02-01 18:51:49
Availability	online
Analysis Status	analysis done

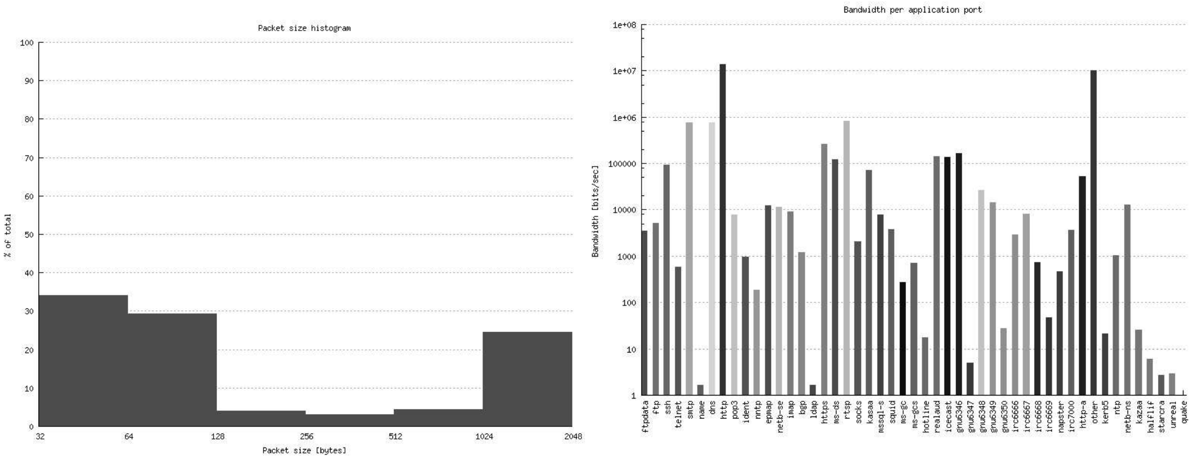
Network Type	WAN
Collector Location	trans-pacific line (USA-Japan)
Traffic Type	operational network, WAN traffic
Link Protocol	Ethernet, 100Mbps
Capture Mode	n/a
Filter Rules	none
Number of Packets	6854781
Trace Anonymisation	IP addresses scrambled using tcpdpriv
Capture Platform	n/a
Data Format	tcpdump
Additional Info	18Mbps CAR was configured on monitored 100Mbps link

Entire traffic rate, averaged over trace duration, in bit/s	27599219
Average packet inter-arrival time in sec	0.000131
Average packet size in bytes	452.9
Average packet arrival rate in pkt/sec	7616.76
Histogram of packet sizes	graph
Bandwidth use per-protocol	graph
Bandwidth use per-application	graph
Series of rates (in bit/s), calculated over 10ms intervals	graph
Series of rates (in bit/s), calculated over 1s intervals	graph
Series of rates (in bit/s), calculated over 3min intervals	graph

Figure 2. Description of exemplary packet trace in MOME database

The analysis results are shown in the lower-right part of Figure 2. Notice that the first group of results corresponds to calculation of the empirical statistical parameters, i.e.: average packet inter-arrival time, average packet size, packet size histogram, average packet rate and average bit rate. The packet size histogram generated by the MOME data analysis workstation for the considered exemplary trace is presented in Figure 3a.

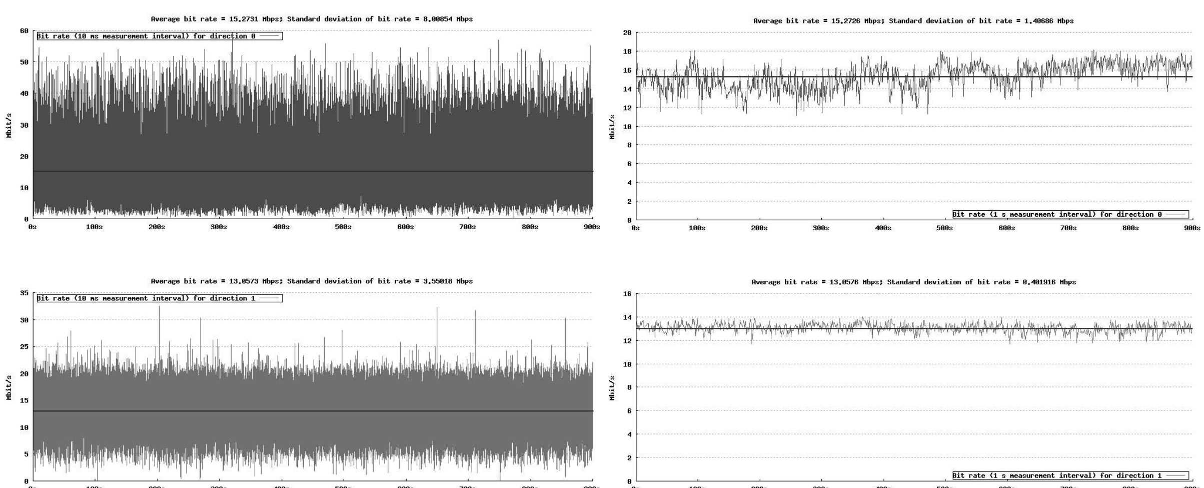
Additional results, which are also accessible in the form of graphs, correspond to the distribution of average bit rate per-transport protocol (TCP, UDP) and per application (identified by the port numbers). These statistical parameters (and graphs) give us basic information about the volume and type of traffic carried on the monitored link (see Figure 3b).



a) b)

Figure 3. Results of MOME data analysis: a) histogram of packet sizes, b) bandwidth usage per-application

Another data analysis result corresponds to the plot of bit rates (calculated as total number of bits transmitted within the time interval, divided by the length of this interval), as a function of time. It gives us information, how the observed traffic fluctuated throughout the trace duration. Notice, that the rates can be calculated over intervals of different length. In MOME database, a few typical values are considered: 10ms (to reveal very-short time scale traffic changes), 1 second (short time-scale), and 3 minutes (medium time-scale). This is illustrated in Figure 4, which shows the time-plots of bit rates produced by MOME data analysis workstation for the discussed exemplary packet trace.



a) b)

Figure 4. Results of MOME data analysis: time-plot of bit rate, calculated in a) 10ms and b) 1s measurement intervals

Remark, that the main goal of MOME Data Analysis Workstation is not to provide access to very sophisticated and advanced analysis tools. The presented results may seem to be rather simple. Certainly, more advanced methods for traffic analysis are known, that provide in-depth insight into the characteristics of captured traffic (see e.g. [3][4]). But, recall that the main goal of MOME is to prepare and store the information, which describes the most important internal features of captured traffic, and which can be helpful for the researchers searching for measurement data appropriate for their needs. Thus, the presented analysis results should be regarded just as additional information, complementary to the description of the measurement scenario and environment included in the MOME meta-data model.

6. Summary

In this paper we described the MOME meta-database, which aims at collecting information about measurement data captured by different research projects. The stored data includes the descriptions of the assumed measurement scenario and environment, as well as the results of data analysis, which may provide additional useful information about the captured traffic. The MOME Data Analysis Workstation allows for performing selected analysis tasks and storing the results directly in the MOME repository. By collecting such information, the MOME project supports the research community with easy access to large amounts of current and appropriately annotated measurement data.

References

[1] IST-MOME web site, *http://www.ist-mome.org*

[2] P. Aranda-Gutierrez et al., „MOME: An advanced measurement meta-repository", in proceedings of 3rd International Workshop on Internet Performance, Simulation, Monitoring and Measurements, IPS-MoMe 2005, Warsaw, March 2005

[3] P. Tran-Gia, N. Vicari (eds.), "Impacts of New Services on the Architecture and Performance of Broadband Networks", COST257 Final Report, compuTEAM, Wuerzburg, 2000

[4] M. Menth (ed.), "Analysis and Design of Advanced Multiservice Networks Supporting Mobility, Multimedia and Interworking", COST279 Midterm Report, Aracne, Roma, 2004

[5] The Internet Traffic Archive web site, *http://ita.ee.lbl.gov/index.html*

[6] The NLANR Network Traffic Packet Header Traces web site, *http://pma.nlanr.net/Traces/*

[7] The NLANR Measurement and Network Analysis web site, *http://watt.nlanr.net/active/maps/ampmap_active.php*

[8] The RIPE NCC Routing Information Service Raw Data, *http://data.ris.ripe.net/*

[9] The MAWI working group traffic archive, *http://mawi.wide.ad.jp/mawi/*

[10] M. Allman, E. Blanton, W. Eddy, "A Scalable System for Sharing Internet Measurements", in Proceedings of Passive and Active Measurements, PAM 2002, Fort Collins, March 2002

[11] The CAIDA "Correlating Heterogeneous Measurement Data to Achieve System-Level Analysis of Internet Traffic Trends" project, *http://www.caida.org/projects/trends/*

[12] The tcpdump web site, *http://www.tcpdump.org/*

[13] C. Schmoll et al., "State of Interoperability", Deliverable D11 of IST MOME, June 2004

[14] The ENDACE web site, *www.endace.com*