Distributed Network Traffic Monitoring and Analysis using Load Balancing Technology

Soon-Hwa Hong, Jae-Young Kim, Bum-Rae Cho
and James W. Hong

DP&NM Lab.
Dept. of Computer Science and Engineering
POSTECH, Pohang Korea
Email: {padosori, jay, brcho, jwkhong}@postech.ac.kr
http://dpnm.postech.ac.kr/
Introduction

Network traffic (text, image, software, audio, video) is increasing continuously both on the Internet and Intranet.

A simple, accurate and efficient network traffic monitoring and analysis is required to understand the current usage as well as to plan for future.

Many shortcomings exist in currently available monitoring systems.

- cannot analyze long-term traffic.
- do not have monitoring capability from multiple network points.
- capture, analysis and presentation all in one machine.
- cannot prevent packet drops from the system overload.

WebTrafMon II attempts to overcome these shortcomings using distributed architecture and load balancing technique.
## Related Work

<table>
<thead>
<tr>
<th>Capture Method</th>
<th>Analysis Method</th>
<th>Analysis Interval</th>
<th>Analysis Scope</th>
<th>Support for Multiple network points</th>
<th>Distributed Architecture</th>
<th>User Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>tcpdump</td>
<td>libpcap</td>
<td>real time</td>
<td>current</td>
<td>layer 7</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>ntop</td>
<td>libpcap</td>
<td>batch</td>
<td>5 second, hourly</td>
<td>layer 7</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>ethereal</td>
<td>libpcap</td>
<td>real time, batch</td>
<td>current, user-specified time</td>
<td>layer 7</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>MRTG</td>
<td>snmp agent</td>
<td>batch</td>
<td>5 minute, hourly, daily, weekly, monthly</td>
<td>layer 2</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>WebTrafMon</td>
<td>libpcap</td>
<td>real time, batch</td>
<td>current, hourly</td>
<td>layer 7</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>WebTrafMon II</td>
<td>libpcap</td>
<td>real time, batch</td>
<td>current, hourly, daily, monthly, yearly</td>
<td>layer 7</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>
WebTrafMon II Requirements

- analyze various types of information: host information, network, transport, and application layer protocols.
- analyze present real-time, hourly, daily, monthly and yearly network traffic data automatically.
- analyze multiple network points traffic.
- no packet drops.
- web-based graphical user interface.
Centralized vs. Distributed

Centralized Traffic Analysis Architecture
- System overload occurs frequently and many packets are dropped
- Cannot support for multiple network points
- Presentation time is slow

Distributed Traffic Analysis Architecture
- Capture, analysis and presentation modules execute on separate machines to minimize system overload
- Can support for multiple network points
- Presentation time is fast
WebTrafMon II: Design

1. multiple network point packet capture and analyze packet header

2. analyze packet header and save into DB and make short term and long term traffic data

3. query to database from user request and give information to user

- probe
  - promiscuous mode packet capture

- analyzer
  - packet header information
  - analyze packet header and save into DB
  - (real-time, hourly, daily, monthly, yearly)
  - make real-time, short term, long term traffic data

- database
  - statistics network traffic data

- web viewer
  - request
  - response

- user
  - distributed environment
Packet Capture Module (Probe)

- Probe captures packet with promiscuous mode, analyzes packet header and saves into log file.
- No packet drops from system overload using an independent packet capture module.

Log File

- Log format
- Log format
- Log format
- Log format
- Log format

Log format: time, length, frame_type, source ip, destination ip, protocol, source port and destination port information
Log Format

- **Log format**:
  - **time**: 4 bytes
  - **length**: 2 bytes
  - **frame_type**: 2 bytes
  - **src_ip**: 4 bytes
  - **dst_ip**: 4 bytes
  - **protocol**: 1 byte
  - **src_port**: 2 bytes
  - **dst_port**: 2 bytes

- **Ethernet**
  - **Dst addr**: 6 bytes
  - **Src addr**: 6 bytes
  - **Type**: 2 bytes
  - **data**: 46-1500 bytes
  - **CRC**: 4 bytes

- **IP**
  - **Src addr**: 4 bytes
  - **Dst addr**: 4 bytes
  - **Protocol**: 1 byte
  - **IP data**: 4 bytes

- **TCP**
  - **src port**: 2 bytes
  - **dst port**: 2 bytes
  - **TCP data**: 4 bytes

- **total Ethernet frame size**: (Ethernet header + data + CRC)
- **frame capture time**
An Analyzer is divided into a Log Transformer module and DB Analyzer module.

- Log Transformer assorts log files into IP-based data, and non IP-based data (e.g., ARP, RARP, IPX).
- Log Transformer saves these assorted data to database.
- DB Analyzer analyzes assorted data in database and makes statistical real-time, hourly, daily, monthly and yearly data.
Data Translation by Analyzer for Long-Term Traffic

Log files

- raw_ip_table
- raw_non_ip_table

hourly

- hourly_data_sent
- hourly_data_received
- hourly_data_exchanged
- hourly_network
- hourly_transport
- hourly_application

daily

- daily_*_table

monthly

- monthly_*_table

yearly

- yearly_*_table
This is a *_network_table example. network_table has IP information from raw_ip_table and ARP and so on information from raw_non_ip_table.
Web Viewer Module Design

Database
- raw_ip, non_ip table
- data_sent table
- data_received table
- data_exchanged table
- network table
- transport table
- application table

Web Viewer

Web Browser (user)

Client

Server

CGI

query
reply
request
response
Implementation

Binary Log File

- packet capture using libpcap
- save packet header information into log file

MySQL Client

- make real-time, hourly, daily, monthly and yearly host, protocol and application information table

MySQL Server

- query
- reply

C-CGI, MySQL Client

- database raw ip, non-ip, table
- data sent table
- data received table
- data exchanged table
- network table
- transport table
- application table

NFS distributed environment

- probe
- web viewer

C-CGI, Apache web-server

http

Internet

user
Web-based User Interface: Main View

- **Traffic History: Hour View**
  - **Monitoring Time:** 2001-04-28, 23
  - **Total number of packets:** 6,420
  - **Total size of packets:** 1,002,683 bytes

- **Host Information**
  - **Data Sent**
  - **Data Received**
  - **Data Exchanged**

- **Protocol Information**
  - **Network Layer**
  - **Transport Layer**
  - **Application Layer**

- **Detailed Views**
  - Hour View
  - Day View
  - Month View
  - Year View

- **Menu**
  - Home
  - Configuration
  - Traffic History
    - Hour View
    - Day View
    - Month View
    - Year View

- **Analysis Time**
  - Time Interval menu

- **Hourly Total Traffic**

Developed by DP&M, POSTECH
Detailed Views

1. Host Information: Data Sent (TOP 10)

2. Host Information: Data Received (TOP 10)

3. Host Information: Data Exchanged (TOP 10)

4. Protocol Information: Network Layer (TOP 10)

5. Protocol Information: Transport Layer (TOP 10)

6. Protocol Information: Application Layer (TOP 10)
Summary & Future Work

- WebTrafMon II overcomes many shortcomings of existing monitoring and analysis systems.
- It can analyze real-time and hourly, daily, monthly and yearly network traffic data.
- With load balancing, independent packet capture prevents packet drops from monitoring system overload.
- WebTrafMon II can analyze multiple network points traffic.

Future work on WebTrafMon II
- More analysis on host and application relationships.
- Adapt to monitor and analyze other types of IP networks (IPoA, IPoWDM, etc.)
- Traffic analysis based on contents (video, audio, etc.)