Introduction lecture

Lecture slides for S-38.3183
Internet traffic measurements and measurement analysis
16.3.2006
Mika Ilvesmäki

Course details

• New course, so everything is done for the first time!
• 4-6 exercises (Matlab etc.)
• Course aim is to give basic knowledge on packet and flow measurements in IP networks
  – Focus is on layers 3 and 4 (IP and TCP)
• Course material
  – Lectures
  – Lecture slides
  – Exercise materials
  – "Chapter 2"
  – Selected scientific articles
• After the course you should
  – Master basic statistical tools
  – Be able to perform traffic analysis of packet and flow phenomena
  – And make basic conclusions
  – Understand different types of measurements

Course contents

• Course material
  – Lecture notes, chapter or two from an (hopefully) upcoming book
  – ~12 lectures
  – Remember to sign via WWWTopi!
• 4-6 exercises (mandatory)
  – Matlab experience required
  – Programming skills recommended
  – In addition to correctness of the answers, the work process influences the grading of the exercises!
• Grading based on final exam. Points gathered from exercises may replace some points in the final exam.
  – Final exam 10.5.2006 9am-12, hall S3
  – Remember to sign up!
Contact information

- Course webpages are the main media for communication
- General: mika.ilvesmaki@netlab.hut.fi
  - Reception on Thursdays (18.3-8.4.2005) after the afternoon lecture for 30 minutes.
- Exercises: Please contact the exercise lecturer
- Other personnel:
  - markus.peuhkuri@netlab.hut.fi
  - marko.luoma@netlab.hut.fi

Why measure?

- To give background to new theories
  - to verify existing theories
  - traffic and network characterization
- To get knowledge of the network status
  - availability
  - use of resources
  - security status
  - network monitoring and control

Who measures?

- Users
  - Application performance monitoring
  - End-to-end performance
- Operators
  - Billing information
  - Performance indicators
    - link utilization, error and loss rates, delays
- Vendors/manufacturers
  - Design improvement

What is there to measure?

- Network events
  - The event itself
    - Count of packets
  - The size or some other quantitative property of the event itself
    - Packet size, flow duration
  - Inter-event relation
    - Frequency of events, the time between two events
- Protocol/Applications behaviour and analysis
  - Requires assembling the packets to messages, content, protocol state etc.
Measurement types

- Mode: active or passive
- Location: single point - multipoint

<table>
<thead>
<tr>
<th>Mode</th>
<th>Location</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active + single point</td>
<td>Passive + multipoint</td>
<td></td>
</tr>
<tr>
<td>Active + multipoint</td>
<td>Passive + single point</td>
<td></td>
</tr>
</tbody>
</table>

Mode: Passive measurements

- No interference to network
- Huge amounts of data
  - Several packets are needed to get accurate information on the network
  - Cf. to sampling. One packet is one sample of the network status, several packets are several samples.
  - Data compression necessary
- Data capture
  - Data copying
  - Passive listening
  - Pass-through

Passive measurement objectives

- Arrival process characterization
  - Packets, flows, applications
- Network status & traffic profiles
- General measures
  - Utilization, traffic trends etc.

Mode: Active measurements

- Measurement probes (packets) injected into the network -> increases the network load and may lead to excess traffic
- Measure for BW capacity, packet delay, packet loss, or RTT
- End-to-end
- Hop-by-Hop (Tunnels)
- Link-by-link
Active measurement objectives

- Current network status
  - Current available bandwidth estimation
  - Current packet loss characteristics
  - Current delay characteristics
  - Current routing status

Type: Measurements at one point

- Measurements done at one point make it possible to analyze
  - Count of events, event InterArrivalTimes, content, volume throughput, round trip times (RTT)
- Analyzing packet contents we can also perform
  - Protocol/Application analysis

Type: Multipoint measurements

- Measurements in two or more points make it possible to analyze and study
  - Delays,
  - Traffic matrices
    - Traffic directionality
  - Clock synchronization
  - Routing behavior

Mode+Type: Active 1-point and multi-point

- Active measurement:
  - Probe sent and response is someway automated from the network by design
- In multi-point active measurements the other end is ready to send response.
Where’s the info on the packet contents?
- Packet header information
  - layers 1 and 2 do not contain any information on packet content
  - layer 3 (IP) identifies the sending source and receiving destination and the upper layer 4 protocol (TCP/UDP)
    - oversimplification: who sends packets where
  - layer 4 (UDP/TCP) identifies the port numbers used at source and destination
    - oversimplification: what application is used
    - source identifies the application that originates the packet and the destination tells us where the packets are headed
- Layers 3 and 4 are the first ones that contain any information on the application that the user is using to create packets in the network.

Grouping packets into flows
- Concept of flow is based on TCP connections
  - Using the TCP protocol, all connections are handled via the SYN and FIN control mechanism. It is therefore possible to watch the traffic on a network, check for SYN and FIN packets and thereby aggregate everything with identical service number, source and destination address etc between the SYN and FIN packet into one “flow”.
    - The strength of this approach is that the detection of beginning and end of a TCP connection based flow is relatively easy.
  - UDP?
- Flow: Packet train model by Jain
  - A packet train is a burst of packets arriving from the same source and heading to the same destination. If the spacing between two packets exceeds some inter-train gap, they are said to belong to different trains.
Flow
• A set of packets that share common information in the header…
  – For instance: srcIP, dstIP, Proto, srcPort, dstPort or parts of these, or any other fields, in the packet headers
• …and appear in the network within a timelimit (timeout)
  – For instance 10 seconds, 60 seconds, 64 seconds, 5, 10 or 15 minutes etc.

Masking
• Filtering packets based on header information
• Network part of SrcIP and/or DstIP
• Host part of SrcIP and/or DstIP
• Protocol (TCP/UDP/other)
• TCP/UDP Sport and/or Dport numbers

Sensitive data in IP&TCP/UDP
• Address fields -> de facto person identification
  – Address space may be determined with checksum –field (TTL has to be guessed)
• Port number may reveal the application used.
• Payload data
  – TCP/UDP checksum (short packets)

Privacy issues
• Traffic contains potentially sensitive information
  – Passwords & Identification data
    • Data privacy
  – Knowledge of existing connections
    • Communication privacy
• Wiretapping and revealing information on parties is strictly controlled
  – Legislation varies from country to country
Privacy protection

- Collect only information you absolutely need (measurement collector activity)
- Sanitize IP addresses (measurement collector activity)
  - Random numbers -> topology lost
  - Lowest order –byte replaced -> protection of single users, preserving routing info
  - Subnetwork and host replaced -> topology preserved
- Encrypt payload (user activity)
  - IPsec, TLS, SSH

“Timescales” of this course

- This course focuses primarily on packet and flow level measurements and analysis