Introduction lecture

Lecture slides for S-38.3183
Internet traffic measurements and measurement analysis
14.3.2007
Mika Ilvesmäki

Contents

- Course arrangements
- Who, what, why measure?
- Active & Passive measurements
- Single-point & Multi-point measurements
- IP (v4 & v6) packet structure
  - TCP and UDP structure
  - packet selection for measurements, masking
- Flow as a measurement concept
- Security & legislation

Course details

- This course can be included into post-graduate studies.
- Relatively new course, so everything is done for the second time!
- 3-5 exercises (Perl, Matlab etc.)
  - Revised from last year.
- Course aims to give basic knowledge on packet and flow measurements and their analysis in IP networks
  - Focus is on OSI layers 3 and 4 (IP and TCP/UDP/Other)
- Course material
  - Lectures
  - Lecture slides
  - Exercise materials
  - ”Chapter 2: Statistics” and possibly ”Chapter 4:Flow analysis”
  - 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 in via WWWTopi!
- 3-5 exercises (compulsory)
  - Weekly returns. No extensions will be granted.
  - Matlab experience required.
  - Programming skills needed (perl, awk, shell-scripts).
  - 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 9.5.2006 9am-12, hall S3
  - Remember to sign up using WWWtopi!
- Feedback: In order to qualify for grading you have to give feedback on the course at http://palaute.ee.hut.fi/

Contact information

- Course webpages are the main media for communication
- General: mika.ilvesmaki@netlab.tkk.fi
  - Reception by appointment.
- Exercises: Please contact the exercise lecturer
- Other personnel:
  - markus.peuhkuri@netlab.hut.fi
  - marko.luoma@netlab.hut.fi

Why measure?

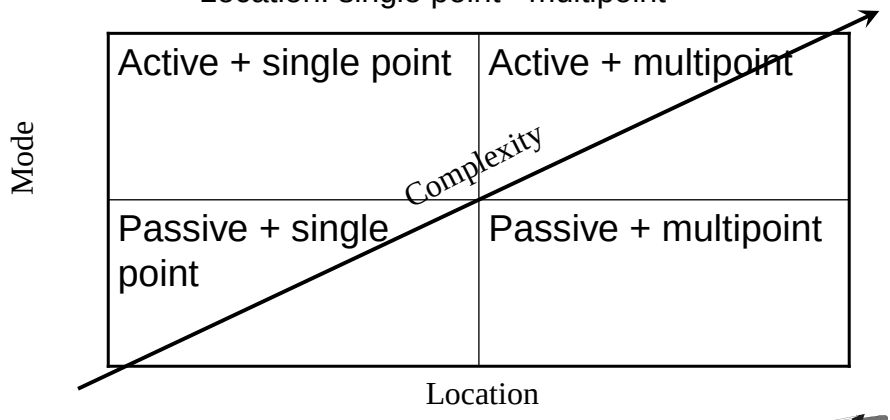
- To give background to new theories
  - to verify existing theories
  - > traffic and network characterization
- To get knowledge of the network status
  - availability
  - service level
  - use of resources
  - security status (anomalies)
  - > network design, monitoring and control

Who measures?

- Users
  - Application performance monitoring
  - End-to-end performance
    - QoS and SLAs
- Operators
  - Billing information
  - Performance indicators for network design and upgrades
    - link utilization, error and loss rates, delays
      - SLAs
- Vendors/manufacturers
  - Design improvement
- Researchers
  - Looking for ways to understand network phenomena better.

Measurement types

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



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, users
- Network status & traffic profiles
- General measures
  - Utilization, traffic trends etc.
  - Protocol shares, user counts

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 bandwdith estimation
  - Current packet loss characteristics
  - Current delay characteristics
  - Current routing status
- SLA measurements

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.

HELSINKI UNIVERSITY OF TECHNOLOGY Mika Ilvesmäki, D.Sc. (Tech.)

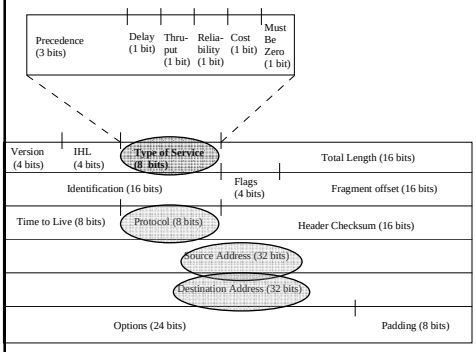
What is there to measure?

- 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.

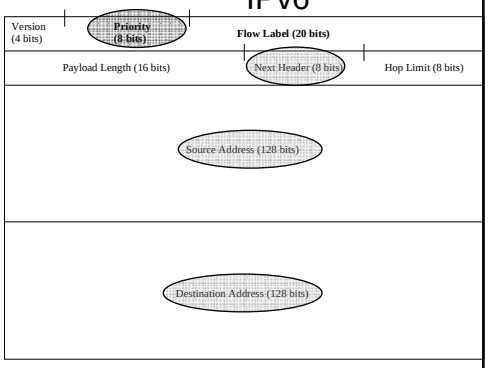
IP-packet structures

Some of the possible aggregation points indicated

IPv4

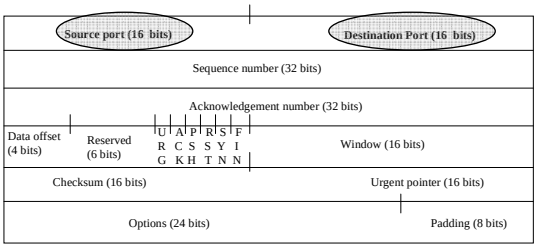


IPv6

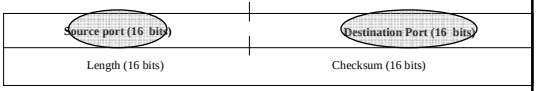


TCP/UDP packet structure

TCP



UDP



Where's the info on the packet contents?

- Packet header information
  - layers 1 and 2 do not contain any (relevant) information on packet content (as far as this course is concerned)
  - 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.
- Layer 3 and 4 info is used as dis/aggregation points

HELSINKI UNIVERSITY OF TECHNOLOGY Mika Ilvesmäki, D.Sc. (Tech.)

Dis/Aggregation

- Dis/Aggregation means the extraction and combination of measured events (or their properties) into new disaggregated or aggregated events.
  - Temporal aggregation: e.g., making timeseries (with equal time intervals) from event based traces
  - Spatial aggregation: e.g., counting TCP and UDP data shares
  - Spatial disaggregation: Extracting all web-server based packets for traffic modeling

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 (or other events) 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

- Enabling traffic dis/aggregation
  - 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