

HELSINKI UNIVERSITY OF TECHNOLOGY

# Flow measurement basics

Lecture slides for S-38.3183  
30.3.2006  
Mika Ilvesmäki

Networking laboratory

HELSINKI UNIVERSITY OF TECHNOLOGY

Mika Ilvesmäki, Lic.Sc. (Tech.)

## Timescales of different events

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

The diagram illustrates the timescales of different events in network measurement. A central circle represents the core measurement level, containing 'flows', 'packets', and 'connections'. Arrows point from this core to four other areas: 'Resource allocation' (top), 'IP applications' (right), 'packet classification' (bottom), and 'slow phenomena' (far right). A vertical dashed line separates the left side (packet/flow level) from the right side (application level). The left side includes 'bits', 'bytes', 'packets', and 'connections', with 'buffering' and 'TCP' also indicated. The right side includes 'IP applications' and 'slow phenomena', with 'traffic classification' (resource based and user based) also indicated.

HELSINKI UNIVERSITY OF TECHNOLOGY

Mika Ilvesmäki, Lic.Sc. (Tech.)

## Contents

- (de)Motivation for flow measurements
- Definitions for flow and related concepts
- Types of flow classification methods
- Flow classifier implementations
- Flow measurements as an aid in understanding network characteristics

HELSINKI UNIVERSITY OF TECHNOLOGY

Mika Ilvesmäki, Lic.Sc. (Tech.)

## Goals of this lecture

- After this lecture you should be able
  - To discuss the pros and cons of flow classification
  - To discuss various concepts related to flow
  - Understand the flow parameters and their effect to flow statistics
  - To discuss the different types of flow classifiers
  - Construct various types of flow classifiers on a functional pseudo-code level
  - To discuss the various applications of flow classification

HELSINKI UNIVERSITY OF TECHNOLOGY Mika Ovaska, Lic.Sc. (Tech.)

## Why flow measurements?

- Packet measurements are not able to give a good "helicopter" view on general application behavior
- Flow classification indicates the traffic characteristics to a degree
  - Use of an application consists of "sessions" where packets are sent and received
  - Flow measurements aim to capture (and analyze) these sessions and their behavior. (The Lunch)
    - While giving up on the knowledge of individual packets within the flow. (Price of The Lunch)
- Depending on the definition of the flow there are almost always less flows than packets observed (memory/storage saved).

HELSINKI UNIVERSITY OF TECHNOLOGY Mika Ovaska, Lic.Sc. (Tech.)

## Flow – concepts

- Definitions:
  - Packets between TCP Syn and TCP Fin
  - Packet train (IEEE JSAC Sept. 1986):
    - The packet train (Flow) model consists of a sequence of packets traveling between a given pair of nodes.
    - Subsequent work has often focused on flows traveling to only one direction
      - [Src:A,Src:B]<->[Src:B,Src:A]
- Networks & routers seldom recognize flows
  - Flow is an abstract network phenomena invented to better model Internet traffic
  - **If (and when) you use flows, be sure to be absolutely precise on the definition**

HELSINKI UNIVERSITY OF TECHNOLOGY Mika Ovaska, Lic.Sc. (Tech.)

## Flow definition for this course

- A flow is (an event that consist of) a set of packets (other events) that are travelling from a source to a destination -> synthetic event
  - 5-tuple flow
    - Fine/High granularity definition -> large number of flows
  - Timeout may vary
    - But is usually 60 or 64 seconds
    - Other work has used timeouts up to several tens of minutes
- Flow definition is always constrained to the measurement/analysis objective

HELSINKI UNIVERSITY OF TECHNOLOGY Mika Ovaska, Lic.Sc. (Tech.)

## Flow granularity

- 5-tuple is the most common
  - Fine granularity, "user/application level"
    - SrcIP, DstIP, Proto, SrcPort, DstPort
- Flows can also be formed based on the network part of the address
  - Identifies flow/traffic directions
  - Additional mask fields help to identify application directionality and spread

HELSINKI UNIVERSITY OF TECHNOLOGY Mika Ovaska, Lic.Sc. (Tech.)

---

## What to measure

- Flow occurrence in timedomain
  - Flow IAT (distributions, time series)
- Flow size
  - Length in packets or bytes, duration in time
    - Flow duration is the actual duration ( $t(\text{pkt}_{\text{last}}) - t(\text{pkt}_{\text{first}})$ )
- Number of flows
  - Significance determined by comparison to other flow counts

HELSINKI UNIVERSITY OF TECHNOLOGY Mika Ovaska, Lic.Sc. (Tech.)

---

## Preconditions for flow – header info

- Preconditions for the flow to exist
  - New fivetuple (always) creates a flow (or a flow candidate)
    - Additional check (with a mask) might be performed to some header field(s) to promote the candidate to an actual flow
      - Sport: Indicating traffic source
      - Dport: Indicating traffic destination
    - A group of flows may be divided into subgroup of flows (with common Sport or Dport etc.)
      - Refer to exercise!

HELSINKI UNIVERSITY OF TECHNOLOGY Mika Ovaska, Lic.Sc. (Tech.)

---

## Preconditions for flow – events

- Flow, a synthetic event, is based on packets (basic events)
- Flows may be created if certain properties of event(s) are fulfilled (within a certain timeframe)
  - Packet count
  - Volume level
  - Packet IAT property
    - Distribution observation
- Event criteria may be combined with header info

HELSINKI UNIVERSITY OF TECHNOLOGY Mika Ovaska, Lic.Sc. (Tech.)

---

## Flow timeout

- As packets enter the network, they are spaced arbitrarily in time by the applications creating them.
  - Flow timeout is the limiting time between two consecutive packets with identical 5-tuples to be associated with the same flow
  - If two packets from the same 5-tuple are not within the timeout limit they belong to different flows
- Flow timeout effects the number of active flows
  - The effect depends up on the application behavior

HELSINKI UNIVERSITY OF TECHNOLOGY

Mika Ovaska, Lic.Sc. (Tech.)

## Effect of flow timeout to flow count

- Different application types produce different amounts of flows depending on the timeout
  - As the timeout approaches the fundamental packet spacing time of the app, every packet creates its own flow
    - Upper limit for flow count -> number of packets
    - Lower limit for flow size -> limited packet size

HELSINKI UNIVERSITY OF TECHNOLOGY

Mika Ovaska, Lic.Sc. (Tech.)

## Implementing a flow classifier

- The number of active flows (at any moment of time) is not known in advance
  - Active flow database needs to be dynamic
- Flows need to be unambiguously identified
  - Otherwise flows/flowdata may be overlapping
  - Need for unique flow id (based on 5-tuple)
- Flows have timeout
  - Needs constant monitoring and updating
    - Latest packet arrival on a flow resets the timeout-monitor
- Flow simulation can be event based but on-line flows need to be monitored continuously

HELSINKI UNIVERSITY OF TECHNOLOGY

Mika Ovaska, Lic.Sc. (Tech.)

## Pseudo code – basic flow classifier

- Hashing/Hash key forming not shown
- Major functionality
  - Packet/Flow classification
  - Flow database maintenance
    - New flows, removed flows
    - Timeout monitoring

```

process Init
  F, D = 0;
  m = (32, 32, 8, 16, 16); % 5-tuple masking
end Init;

process Flow_Classification
  packet p arrives--
    F = F;
    if ∃ f ∈ F such that % packet belongs to existing flow
      pf ∈ f and t - t_f < τ and M(m, p) = M(m, pf);
    then
      F = F \ D; % remove flows to be deleted
      t_f = t;
    else
      F = (F ∪ f_p) \ D; % remove flows to be deleted and add new flow
      t_f = t;
    fi;
    D = { f | t - t_f ≥ τ }; % check for timeout
  end Flow_Classification;
  
```

HELSINKI UNIVERSITY OF TECHNOLOGY

Mika Ovaska, Lic.Sc. (Tech.)

## Packet count flow classifier

- One common flow precondition is the packet count
  - The number of packets (within a certain time) is observed on a *flow candidate*
    - No time limit means the candidate waits for the packet count threshold forever.
  - As the packet count (and time) precondition is met the flow candidate is promoted to an active flow
    - A sort of a high pass filter for packet streams
  - If precondition is not met, the flow candidate times out

HELSINKI UNIVERSITY OF TECHNOLOGY Mika Suvarinki, Lic.Sc. (Tech.)

## Implementing a packet count classifier

- Same requirements as for the regular flow classifier
- In addition, there exists a database for candidate flows
  - The preconditions have to be monitored and candidate database should be maintained
  - The more complex the preconditions criteria the more complex the database maintenance

HELSINKI UNIVERSITY OF TECHNOLOGY Mika Suvarinki, Lic.Sc. (Tech.)

## Pseudo code – packet count classifier

- Candidate flow database
- Active flow database
- Thx to Jouni Karvo for the pseudo code

```

process packet_count
  C(f, x, t_p) = ∅;
  F(f, t_p) = ∅;
  packet p arrives →
  if
    ∃ f_p ∈ C: p_f = C(f_p) ∧ C(f_p, x) < X → % check and update candidates
      C(f_p, x) = C(f_p, x) + 1; C(f_p, t_p) = t;
    ∃ f_p ∈ C: p_f = C(f_p) ∧ C(f_p, x) ≥ X → % candidates to active flows
      F = F ∪ f_p; C = C \ f_p;
    ∃ f_p ∈ C: p_f ≠ C(f_p) ∧ ∃ f_p ∈ F: p_f ≠ F(f_p) → % new candidates
      C = C ∪ f_p; C(f_p, x) = 1; C(f_p, t_p) = t;
    ∃ f_p ∈ F: p_f = F(f_p) → % update active flow status
      F(f_p, t) = t;
  fi;
  do
    ∃ f_p ∈ C: t - C(f_p, t_p) > T_C → % remove timed out candidates
      C = C \ f_p;
  od;
  do
    ∃ f_p ∈ F: t - F(f_p, t_p) > T_F → % remove timed out actives
      F = F \ f_p;
  od;
end packet_count;
  
```

HELSINKI UNIVERSITY OF TECHNOLOGY Mika Suvarinki, Lic.Sc. (Tech.)


## Variations of packet count classifier

- Packet count threshold and/or other parameter values are modified according to network/router status
  - Feedback loop, classifier tries to effect the network/router status (adjustable high pass filter)
  - Available resources in the router
    - Buffer space, instantaneous delay etc.
  - Available flow space
  - Available BW on the link

HELSINKI UNIVERSITY OF TECHNOLOGY Mika Suvarinki, Lic.Sc. (Tech.)

## Flow classification as a feedback process in the network

- Flow classification can be used to control the use of network resources
  - Classifier controls the network
- The control criteria and controlled resources are practically limitless
  - Although limits exist to which are feasible





HELSINKI UNIVERSITY OF TECHNOLOGY

Mika Suominen, Lic.Sc. (Tech.)

## Characterizing traffic based on flow data

- What flow data tells about traffic?
  - Everything is related to flow parameters
    - Granularity and timeout
  - High flow count indicates bursty (frequently ending and restarting) traffic or a lot of users
    - Bursty as compared to the timeout value
  - Low flow count indicates stability (long lasting flows) or few users
  - Flow volume, IAT distributions etc.






HELSINKI UNIVERSITY OF TECHNOLOGY

Mika Suominen, Lic.Sc. (Tech.)

## Limitations of flow data based characterization

- Within the timeout limit, there is no indication of the traffic variation
- Long (time)scale phenomena may be hidden behind short timeout limits
  - What traffic characteristics are revealed with 64 second flow timeout? And what if the flow timeout is 48 hours?
- Packet data often gives additional knowledge of the traffic characteristics






HELSINKI UNIVERSITY OF TECHNOLOGY

Mika Suominen, Lic.Sc. (Tech.)

## Off-line Flow analysis

- Flow analysis refers to a process that is performed off-line
  - With adequate resources at hand
  - Flow parameters may be extreme (infinite timeouts etc.) and they may be freely varied
    - Flow analysis takes as input a packet trace file
  - Goal is to gain knowledge of the network and its traffic characteristics for future actions





HELSINKI UNIVERSITY OF TECHNOLOGY

Mika Suominen, Lic.Sc. (Tech.)

## On-line Flow classification

- Flow classification refers to a process that is performed on-line and real-time in a network
  - All available resources are limited
    - Time, processing capacity, storage capacity, etc.
    - Future resource usage is also unpredictable
  - All actions taken and processes started have to be preplanned and designed for quick execution
    - Software and hardware design problems
  - Flow classifier takes as input one single packet at a time





## Flow measurement problems

- Why is per-flow measurement hard?
  - Keeping per flow state is not sensible because of the high cost of maintaining data-structures.
  - Majority of the packets belong to large flows, yet a majority of the flows are small.
  - Several definitions exist for the flow.
  - Worst-case behaviour of data-structures cannot be amortized due to the real time nature of the application.



## Summary of flow measurements

- Flow concepts & definitions
- Motivation for flow analysis
- Applications of flow analysis and flow classification
- Pros & Cons

