Combined (packet & flow) measurements

Lecture slides for S-38.3183

30.3.2006

Mika Ilvesmäki

Contents

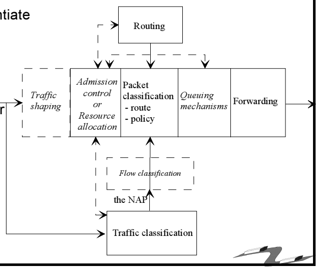
- Meaning of packets and flows
  - presence
  - burstiness
- Means to combine packet/flow data
  - basic measures
  - distributions
  - combinations
- Case: packet & flow counts combined
  - goal
  - tools
  - analysis
  - results

Goals of this lecture

- This is a case lecture
  - Its main purpose is to raise questions
    - It can also produce ideas for further research
  - This lecture documents already finished research and aims to (somewhat) enlight the process that lead to it.
  - This lecture also lets you reflect on the previous lectures and materials. To assess what you have learned and what not…

Motivation

- Measure the network
- Find out application characteristics
  - Use this info to classify/differentiate traffic
  - Find out if applications differ in behavior
  - Develop methods to detect abnormal application beahavior
    - Security threats
    - Network anomalies



Aim for simplicity

- In networks we measure events (packets)
  - Packet: IAT, Length, # of packets
- And we can regroup packets into flows
  - Flow: IAT, length, # of flows, volume
- The most simple statistics are the counts of packets and flows

Packet count

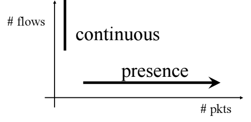
- Packet count indicates application presence
  - More packets, more presence
  - Less packets, less presence
    - Please note, presence is very vaguely defined concept
- Observing merely packets does not give reliable info on
  - Popularity (a lot of users), although high packet count suggests popularity
  - Application behavior (burstiness)

Flow count

- Flow count indicates application behavior and popularity
  - More flows, more bursts (or more users)
  - Less flows, less bursts (or less users)
- Observing flow count does not give reliable info
  - Whether the application is used by large number of users or if it is behaving bursty (or both)
  - On how much the application sends packets

Packets and flows as characteristic measures

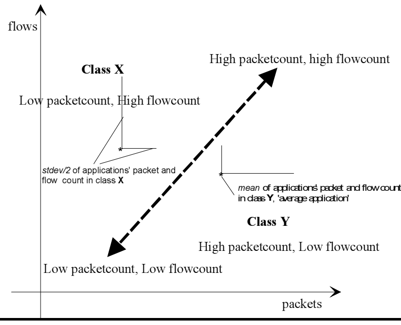
- Packet count indicates the overall application presence
- Flow count gives indication how the application behaves (bursty or continuous)
  - Application behavior is in relation to the value of the flow timeout!



Flow data processing

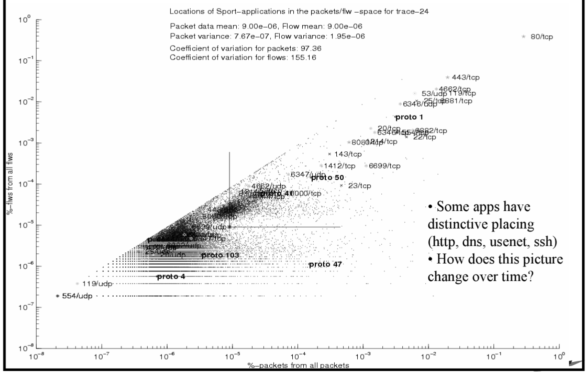
- 5-tuple flow data
  - Sadr,Dadr,Proto(TCP/UDP),Sport,Dport and 64 second timeout
- Group (sum, aggregate) packet and flow counts per existing TCP/UDP Source port
  - The TCP/UDP source port identifies the application
    - with adequate accuracy…
    - For every flow ***f*** with ***Sport*** add the #pkts, #flws and bytevolume -> List of ***Sport***s with #pkts, #flws and bytevolume
    - Normalize #pkts,#flws and bytevolume against Σpkts, Σflws and Σbytevolume and plot to 2d scatter plot

Scatter plot of packets and flows

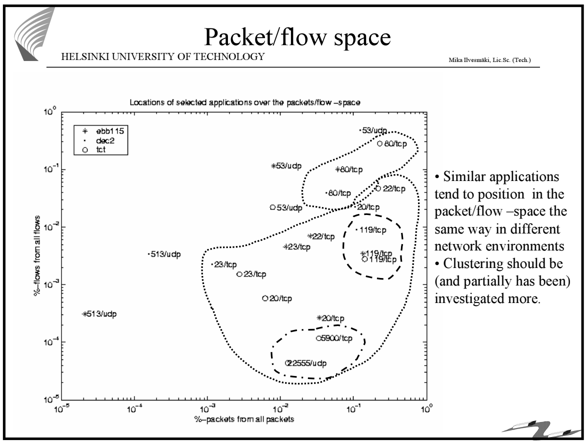

flows
Class X
High packetcount, high flowcount
Low packetcount, High flowcount
stdev/2 of applications' packet and flow count in class **X**
mean of applications' packet and flowcount in class **Y**, 'average application'
Class Y
High packetcount, Low flowcount
Low packetcount, Low flowcount
packets

Results

- Academic achievements: 1 doctoral thesis, approximately 10-15 publications
- Scientific(?!) results
  - Applications cluster in pkt-flw –space
    - The cause of this is still unclear
    - One algorithm developed to utilize clustering
  - Flow count changes differently depending on application nature
    - Partly because different applications produce packets differently
      - TCP might filter out "true" application packet process
      - UDP apps show true (or closer to) characteristics



- Some apps have distinctive placing (http, dns, usenet, ssh)
- How does this picture change over time?

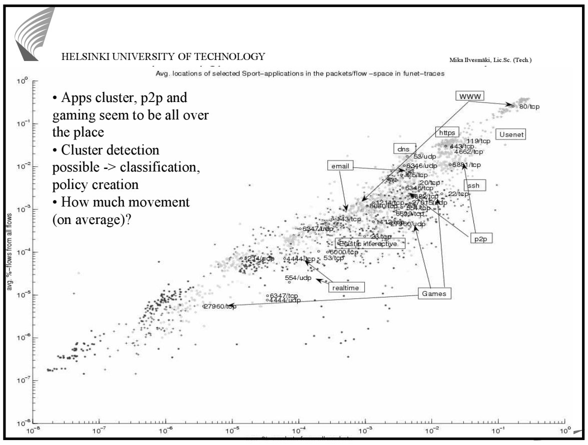


Packet/flow space

HELSINKI UNIVERSITY OF TECHNOLOGY

Mika Ilvesmäki, Lic.Sc. (Tech.)

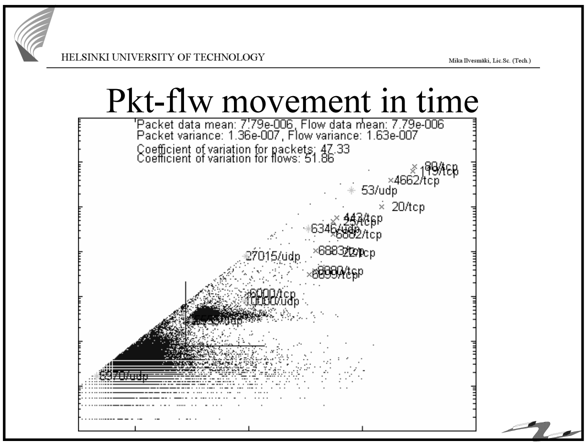
- Similar applications tend to position in the packet/flow –space the same way in different network environments
- Clustering should be (and partially has been) investigated more.



HELSINKI UNIVERSITY OF TECHNOLOGY

Mika Ilvesmäki, Lic.Sc. (Tech.)

- Apps cluster, p2p and gaming seem to be all over the place
- Cluster detection possible -> classification, policy creation
- How much movement (on average)?

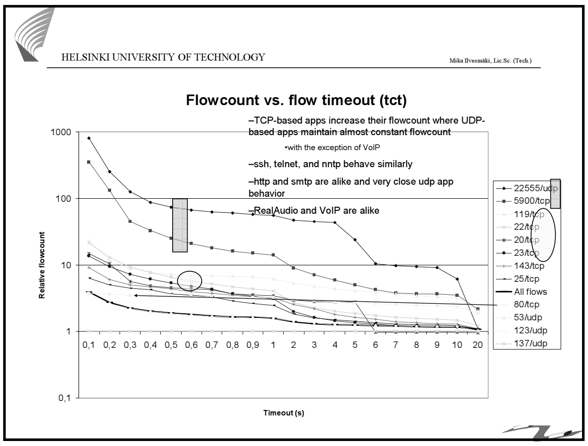


HELSINKI UNIVERSITY OF TECHNOLOGY

Mika Ilvesmäki, Lic.Sc. (Tech.)

Pkt-flw movement in time

Packet data mean: 7.79e-006, Flow data mean: 7.79e-006
Packet variance: 1.36e-007, Flow variance: 1.63e-007
Coefficient of variation for packets: 47.33
Coefficient of variation for flows: 51.86



HELSINKI UNIVERSITY OF TECHNOLOGY

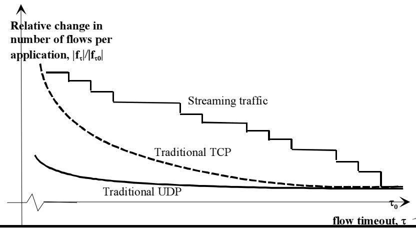
Mika Ilvesmäki, Lic.Sc. (Tech.)

Flowcount vs. flow timeout (tct)

- –TCP-based apps increase their flowcount where UDP-based apps maintain almost constant flowcount
  - •with the exception of VoIP
- –ssh, telnet, and nntp behave similarly
- –http and smtp are alike and very close udp app behavior
- –RealAudio and VoIP are alike

Effect of flow timeout to flow count

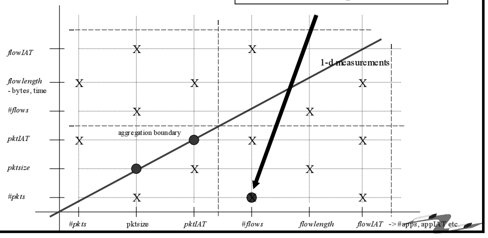
- Different application types produce different amounts of flows depending on the timeout
  - As the timeout approaches the fundamental packet spacing time of the app, every packet is created its own flow
    - Upper limit for flow count -> number of packets
    - Lower limit for flow size -> limited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



Other measurement options in 2D

- Other options ('X's in the figure) for combining measured properties

Our choice: #pkts, #flows



Summary on PacketFlow-measurements

- Application behavior may be characterized (and subsequently classified) by combining packet and flow statistics
  - Packet and flow counts are enough
- This relative application behavior in packet-flow -space seems to remain the same over time