

ELEC-E7130 Internet Traffic Measurements and Analysis

Course introduction and requirements

Markus Peuhkuri Samuli Aalto Juho Kaivosoja

2022-09-07

ELEC-E7130 Internet Traffic Measurements and Analysis

Who we are (and is Zoom recording on)?



Figure 1: Presemo link <https://presemo.aalto.fi/elece7130/>

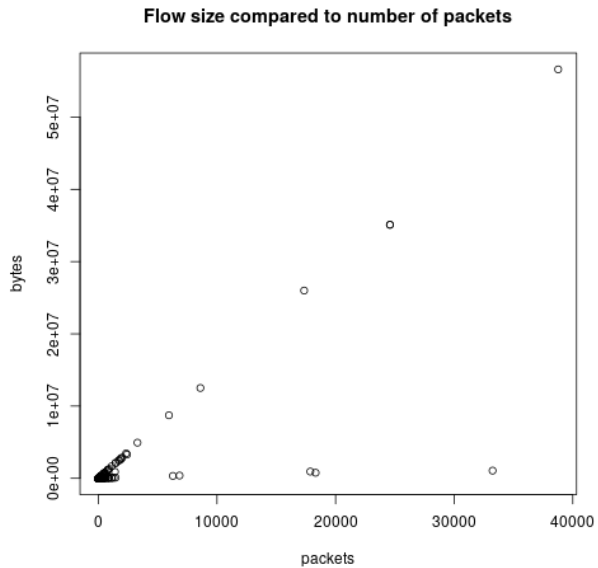
Why measure

- Network Operators
 - SLA monitoring, fault location, anomalies
 - capacity planning, service development
- Users
 - monitor performance
 - check that they receive what they are paid for
- Vendors
 - improve protocol and device design
 - operating practises
- Government
- Law enforcement
- Researchers

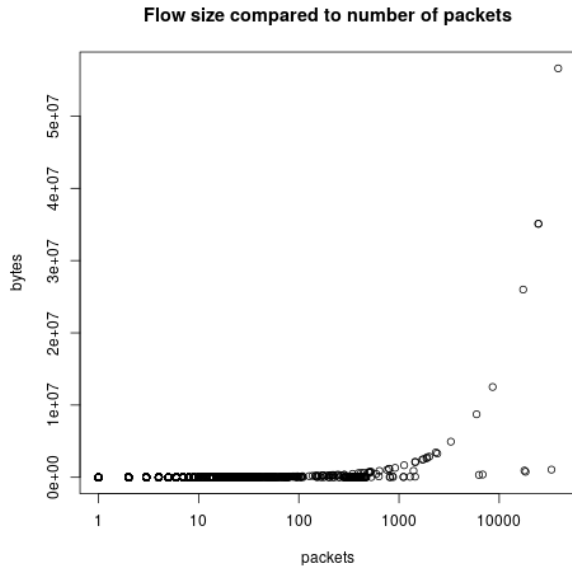
Why analyse

- A single reading from a measurement is next to useless
- Data collected from number of users, devices have already outnumbered humans
- Understanding characteristics of data: if is it
 - random
 - structured
- Big Data, Machine Learning, Artificial Intelligence

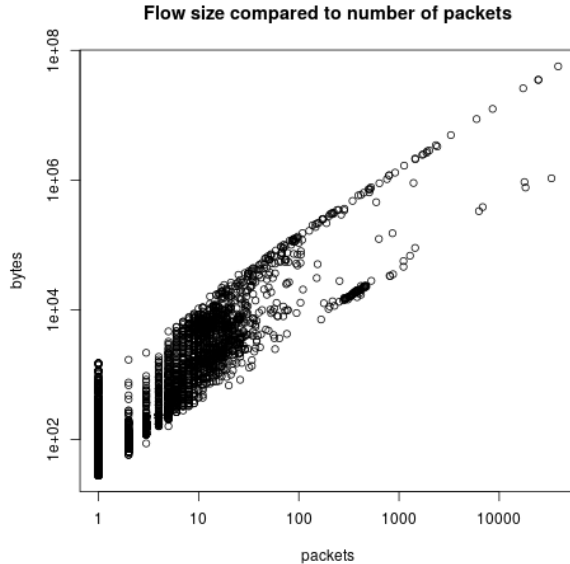
How to visualise findings



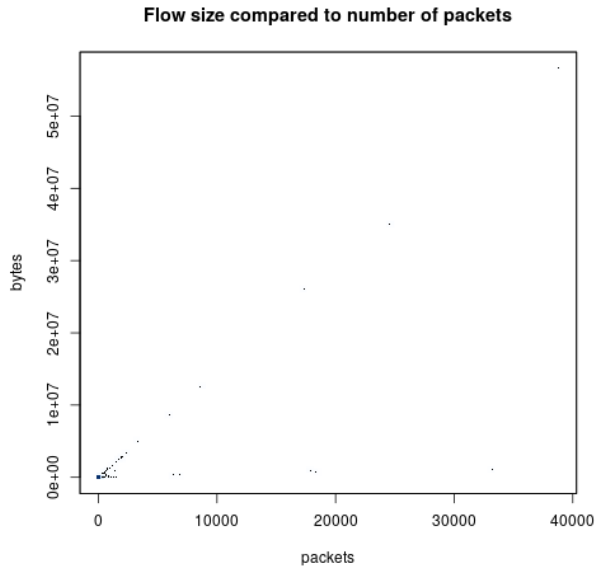
How to visualise findings



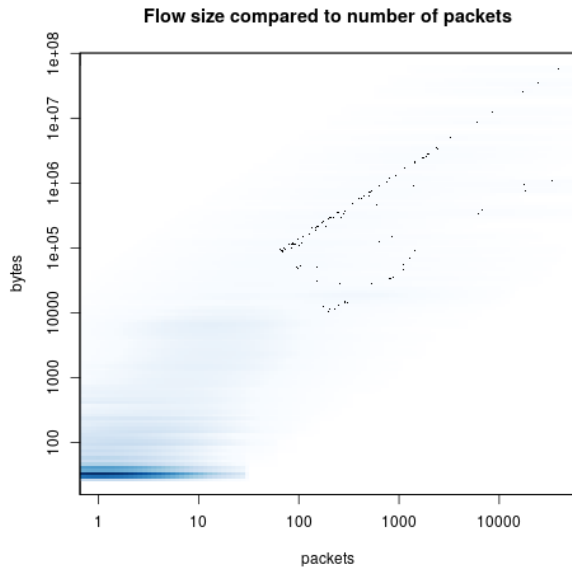
How to visualise findings



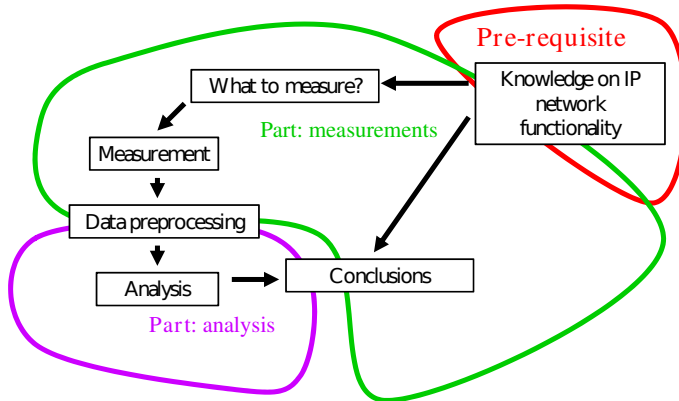
How to visualise findings



How to visualise findings



Network measurement process



Objectives

After the course, you...

1. are able to measure and analyse basic properties of network traffic and draw conclusions on the results
2. are able to apply statistical methods in processing, analysing, and presenting the measurement data; also able to critically evaluate the applicability of the methods
3. understand the technical and legal issues related to network measurements
4. are familiar with methods and tools related to network traffic measurement and analysis

Prerequisites

It's easier to pass the course if you already know:

- Basic knowledge of IP networks. ELEC-C7240 (or equivalent) recommended.
- First course in probability and statistics (MS-A050x)
- Linux command line basics
 - awk, sed, tr, grep, cut, bash
- (Script) programming to make measurements and pre-process data
 - **python**, perl, javascript, java, ruby, php, C++, C#, go, bash, . . .
- Statistical software like **python** (with numpy, pandas and matplotlib) or **R** for analysis
 - Other options include Tableau, matlab, Google Data Studio. . .

Course personnel can best support the **Linux-python-R** toolchain, but you are free to choose the tools you like best. Try Linux on VirtualBox on Windows or OS X if you are unfamiliar with it.

How to pass the course?

- Enroll to the course. Everyone have been successful?
 - Master “Internet Traffic Measurements and Analysis” topics
 - Lectures on Wednesday mornings (non-mandatory but also non-useless)
 - Eight exercise assignments are mandatory and will give
 - Max 40 points
 - Acquire much of skills needed for final assignment
 - If you are not able to make to a some exercise event for some reason, a small extra work is required (actual assignment needs to be returned within time; extra work by December 15th)
 - Mandatory exercise/help events on Thursdays (two-hour slots, 1st period every week) and Mondays (two-hour slots, 2nd period), indicated in schedule
 - Final assignment max 60 points – you need a passing grade (minimum points)
-

(Almost) Weekly exercises (8 instances)

- Introduction on Thursday: initial group discussion and review
- Two (or three) sessions per day: groups opened on 12th September noon
- Dead-line on Wednesday two weeks later 22:00
 - Late return: max 1 point
- Return via MyCourses
 - If you find an error in your submission after dead-line, do *NOT* resubmit the fixed version before receiving acknowledgment from course staff. If you do, **your submission is seen as late.**
- Review on Thursday with discussion and comments
- These are **mandatory**, with option to replace no-show with additional report of an assigned subject (1-2 pages)

Access to weekly exercises

- Will be carried out as on-site sessions in Y342a room
- Course staff will give introduction and available for helping you out
- Remote access to classroom computers available
- Help via course Zulip after sessions

Options for running experiments

- Your own computer
 - Linux recommended
 - Windows users: run virtual Linux, WSL2 is *reported* to work
 - MacOS and *BSD operating systems: beware of different command line usage
- Aalto Virtual desktop <https://vdi.aalto.fi>
 - No heavy computation on virtual hosts
 - Provides full desktop via browser or VMWare Horizon Application
- Aalto Linux servers: `kosh.aalto.fi` and `lyta.aalto.fi` for lightweight processes, `brute.aalto.fi` and `force.aalto.fi` for heavy computation
- Aalto Linux classroom computers
 - Can be accessed with `ssh` via Linux servers or from VDI

Options for running experiments II

- Do not access remotely if there is class on-going. Check from <https://computers.aalto.fi> and <https://booking.aalto.fi>
- Computer names:
<https://www.aalto.fi/en/services/linux-computer-names-in-it-classrooms>

Final Assignment

- Two parts
 1. ready dataset given to analyse
 2. collect your own dataset and analyse it
- Analyse and make a *clear* report. All work must be individual!
- Dead-line by 2nd December noon (2022-12-02T12:00 Finnish time)
 - Late submission gives grade 1 at best; **Return MVR early**, do not resubmit fix after DL (unless agreed with staff)¹
- Review discussion before and on **Monday** 2021-11-28 – you should know how to complete the assignment at this state
 - **Mandatory** event: if you cannot make there for some reason, contact course staff well before dead-line.

¹MVR=Minimum Viable Report

Where to get help to pass the course?

- Exercise sessions on
 - Thursdays 8-16, 2-hour sessions (2022-09-15 – 2022-10-27)
 - Mondays (mostly) 8-16, 2-hour sessions (2021-10-24 – 2021-11-28),
- Discussions on Zulip <https://elec-e7130-2022.zulip.aalto.fi>
 - Both peer support and support by course staff
- Peer support is encouraged but submissions must be individual
 - Plagiarism is **very obvious** when multiple people report the same graphs although data has been different.

Material

- Lecture notes by Markus Peuhkuri
- Slides and extra material provided by lecturers
- Books: (can be found from Aalto library, some as ebook)
 - Data Analysis:
 - David S. Moore and George P. McCabe, *Introduction to the Practice of Statistics*, 5th Edition, W.H. Freeman & Co., 2006 -> Chapters 1,2
 - Sampling and experimental design:
 - David S. Moore and George P. McCabe -> Chapters 3,5
 - Probability models and measurements:
 - Sheldon M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*, 5th Edition, Elsevier, 2014
 - Mark Crovella and Balachander Krishnamurthy, *Internet Measurement: Infrastructure, Traffic, and Applications*, John Wiley & Sons, 2006

Material II

- Stochastic processes in network measurements:
 - Mark Crovella and Balachander Krishnamurthy (above)

Personnel

- Lecturers
 - Markus Peuhkuri markus.peuhkuri@aalto.fi
 - Samuli Aalto
 - Juho Kaivosoja
- Assistants
 - César Olvera Espinosa
 - Markus Holmström
- Best way is to reach via course Zulip
 - you can ping staff with @staff on Zulip in an urgent case. We do monitor discussion quite regularly anyway over working hours, so save it for real need
 - Or by email if problems with Zulip

Just one more thing

Remember:

Ask help before giving up!

Co-operation and helping encouraged but do not plagiarise!