



Aalto University

# ELEC-E7130 Internet Traffic Measurements and Analysis

Course introduction and requirements

Markus Peuhkuri   Esa Hyytiä   Juho Kaivosoja

2023-09-06

# ELEC-E7130 Internet Traffic Measurements and Analysis

# Who we are (and is Zoom recording on)?



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

# What is this course about?

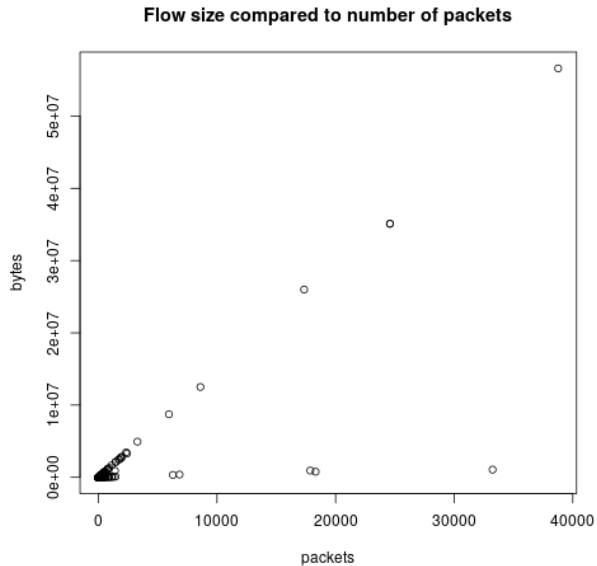
# Why measure

- 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

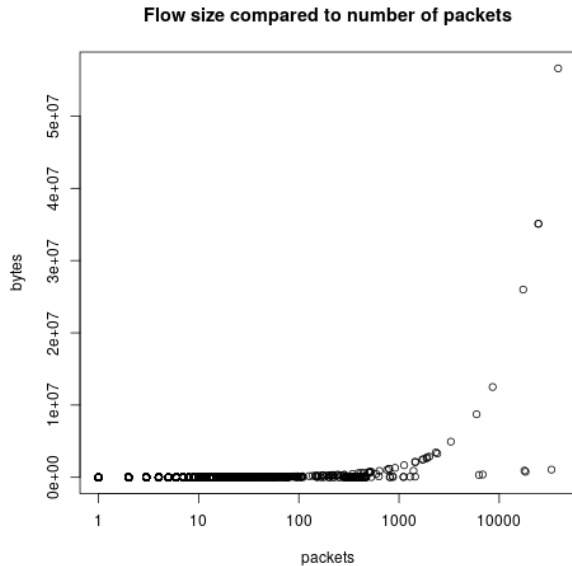
# Why analyse

- 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

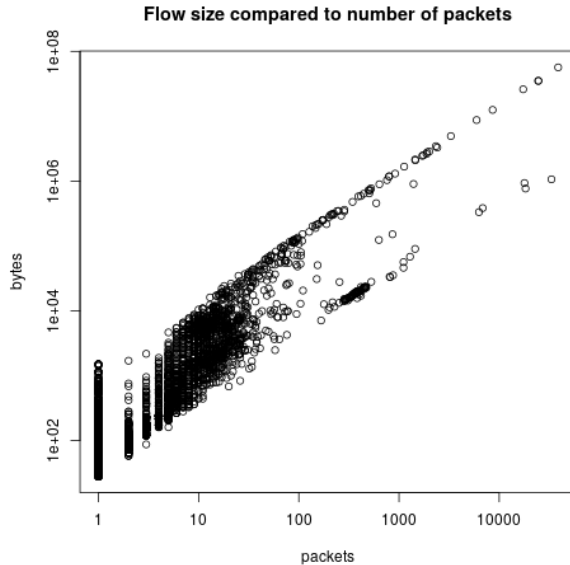


# How to visualise findings

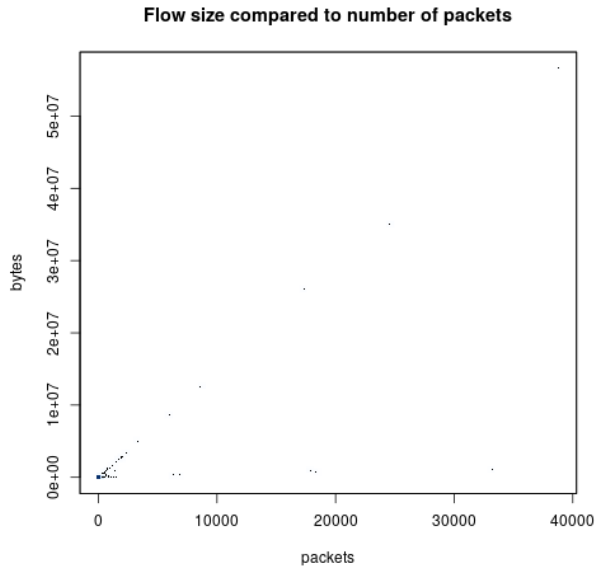




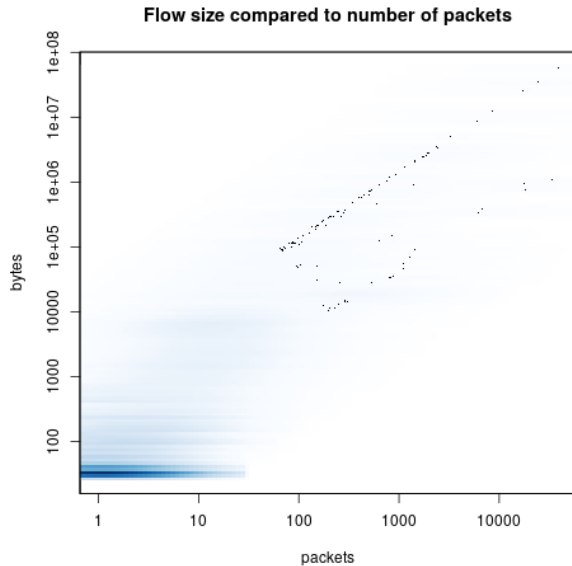
# How to visualise findings



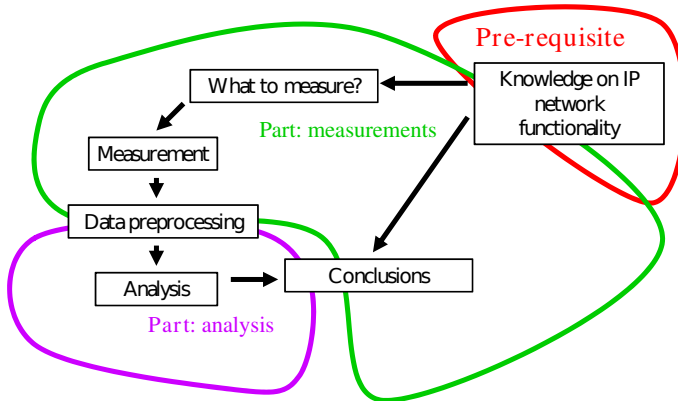
## How to visualise findings



## How to visualise findings



# Network measurement process



# The course: how to complete

# 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.

# 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 weekly exercise/help events as indicated in schedule
- Final assignment max 60 points – you need a passing grade (minimum points)



## (Almost) Weekly assignments (8 instances)

- Introduction on Thursdays (1st and 2nd period) or Mondays (2nd period)
- Dead-line on Wednesday two weeks later at 22:00
  - Late return: points cut half
  - **Return MVR<sup>1</sup> early**
- 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

---

<sup>1</sup>MVR=Minimum Viable Report

# Access to weekly exercises

- Select one of three two-hour sessions
  - Initial questionnaire on preference open until Friday noon
  - Group enrollment opened on Mon 11th September noon (select one)
- Course staff will give introduction, demos, and is available for helping you out
- Check assignment and possible intro video before session so you can ask questions right away
- Remote access to classroom computers available
- Help via course Zulip between sessions
- These sessions are **mandatory**, with option to replace no-show with additional report of an assigned subject (1-2 pages)

# Options for running experiments

- Do I need dedicated computer for the course?
  - If a Linux is your daily driver, you are all set!
  - If you run Windows on your computer:
    - Linux on VirtualBox
    - Windows Subsystem for Linux
  - macOS with command line (e.g. Terminal or iTerm) is good for majority of tasks as well as other \*BSD UNIX systems
    - differences in some command usage and options
- Aalto Virtual desktop <https://vdi.aalto.fi>
  - No heavy computation on virtual hosts
  - Provides full desktop via browser or VMWare Horizon Application

# Options for running experiments II

- 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: `something.org.aalto.fi`
  - Can be accessed with `ssh` via Linux servers or from VDI, e.g. `ssh faure.org.aalto.fi`
  - Do not access remotely if there is class on-going. Check from <https://booking.aalto.fi>
  - Computer names:  
<https://www.aalto.fi/en/services/linux-computer-names-in-it-classrooms>, add `.org.aalto.fi` for a DNS name.
- Last resort: borrow computer from DICE Technical Services
  - Contact course staff for details


# Final Assignment

- Two parts
  1. ready dataset given to analyse
  2. collect your own dataset (starting from A3) and analyse it
- Analyse and make a *clear* report. All work must be individual!
- Dead-line by 4th December **noon** (2023-12-04T12:00 Finnish time)
  - Late submission gives grade 1 at best; **Return MVR early**, do not resubmit fix after DL (unless agreed with staff)
- Review discussion before and on **Monday** 2023-11-27 – 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.

# Where to get help to pass the course?

- Exercise sessions: check assignment before session and ask any questions
- Discussions on Zulip <https://elec-e7130-2023.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 e.g. multiple people report the same graphs although data has been different.
- From AI?

## Want to use AI (for other than in Assignment 8)?

 Generative AI is experimental

rm with confirmation



To use the `rm` command with confirmation, you can use the `-f` option. The syntax is:

```
rm -f <file>
```

Use code with caution. [Learn more](#)



where:

- `<file>` is the file you want to delete

If you use the `-f` option, you will be prompted to confirm before the file is deleted.



# Policy for use of large language models and other AI tools

- See *Guidance for the use of artificial intelligence in teaching and learning at Aalto University*
- AI is just one of *\_(mis)\_* information sources:
  - You are responsible for the correctness of information
  - Sources and tools must be cited and/or indicated
- Inappropriate use of AI tools will be handled as cheating
  - e.g. invalidation of a part or the whole course
  - See *Aalto Code of Academic Integrity*
- So, *do your own research!*



## Bonus points for report quality? Yes! Two of them!

Mandatory criteria to receive the first bonus point (all must be met).

- Student name, id and the assignment information in the cover (or the first) page.
- Structure follows assignment and it is easy to follow.
  - Any subtask without answer is clearly marked missing.
- Tables and figures are correctly sized, labeled correctly and referenced in text.
- References (if any) cited correctly.

Additional criteria towards the second bonus point.

- Well written text:
  - No major misspellings or grammatical mistakes.
  - Has complete sentences with correct punctuation.
- Avoids unneeded duplication of information.
- Colors, markers, axis and labels accessible.

# Material

- Lecture notes by Markus Peuhkuri
- Slides and extra material provided by lecturers
- Books: (can be found from Aalto library, some as ebook)
  - Gareth James, Daniela Witten, Trevor Hastie, Rob Tibshirani, Jonothan Taylor. *An Introduction to Statistical Learning*, Springer 2021, 2023 (PDF download from <https://www.statlearning.com/>)
  - 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:

## Material II

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

# Personnel

- Lecturers
  - Markus Peuhkuri markus.peuhkuri@aalto.fi
  - Esa Hyytiä
  - Juho Kaivosoja
- Assistants
  - Yu Fu
  - Konstantin Mishin
  - Matias Varonen
- 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!