Invited Talk at Copenhagen Business School

Cloud and Fog:
How and Where is My Data Flowing?
Obtaining Insights into Data Privacy in Today's Applications

Thomas Dreibholz (托马斯博士)
Simula Metropolitan Centre for Digital Engineering
dreibh@simula.no

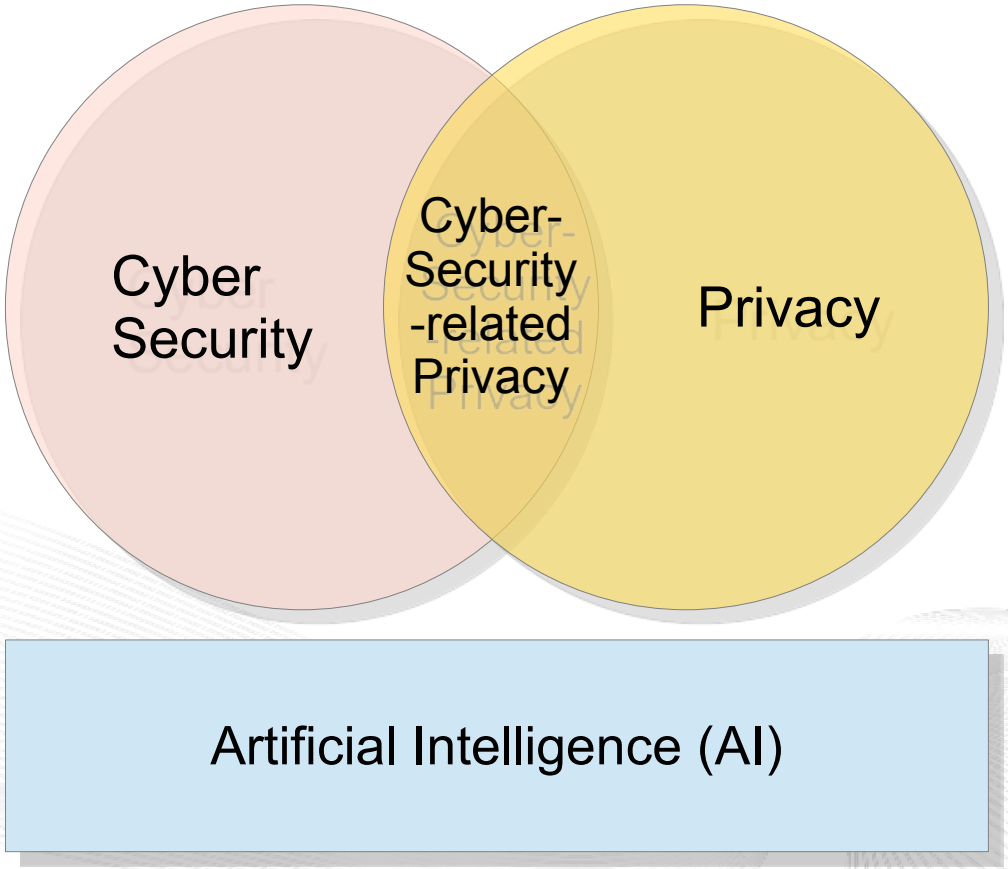
April 26, 2024

Contents

About the Presenter

- Thomas Dreibholz
  - Chief Research Engineer at SimulaMet in Oslo
  - Habilitation in Computer Science in 2012
  - Ph.D. in Computer Science in 2007
  - M.Sc (Diplom) in Computer Science in 2001
  - Experience with Internet protocols since ca. 1996
  - Working with Linux systems since 1994
  - Open Source software development
- Website: https://www.nntb.no/~dreibh/

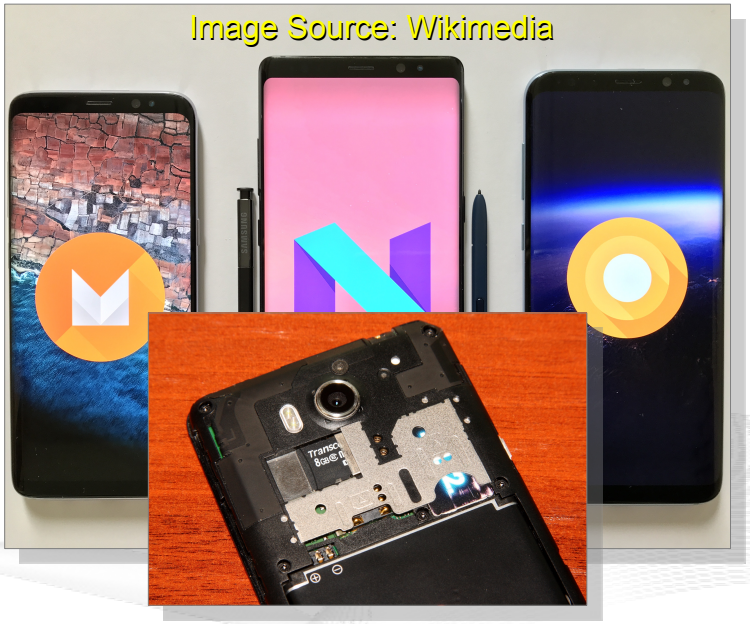
Scope



- Today's talk: Networking
- Networking contains privacy-relevant user data
  - Can be processed by AI to extract user profiles, etc.
- How does networking work?
- What are the issues?
- How to make improvements?

Trend: Smartphones and Cloud Computing

- Smartphone
  - Storage space is small (or expensive)
  - Hardly extensible (e.g. by SD card slot)
- Cloud connectivity
  - **Storage space**
    - Pictures, videos, music, maps, ...
    - Documents
  - **Applications**
    - Computation-intensive applications in the cloud (e.g. voice recognition, AI/ML, ...)



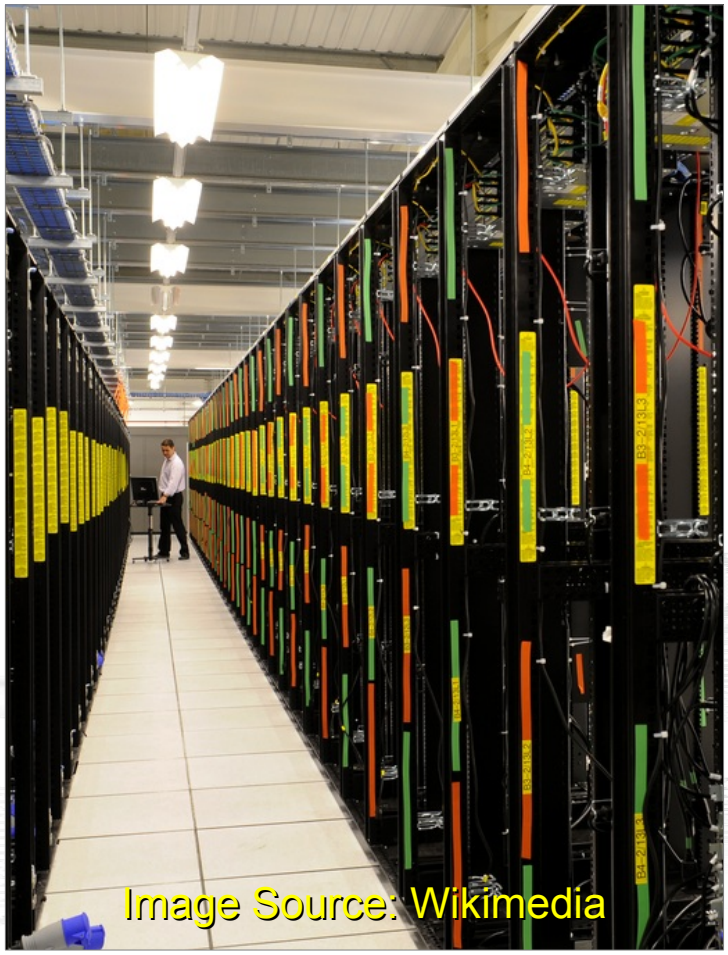
"Cloud" is an integral part of today's smartphones!

What is the “Cloud”?

Image Source:
CompuServe advertisement 1988
Internet Archive

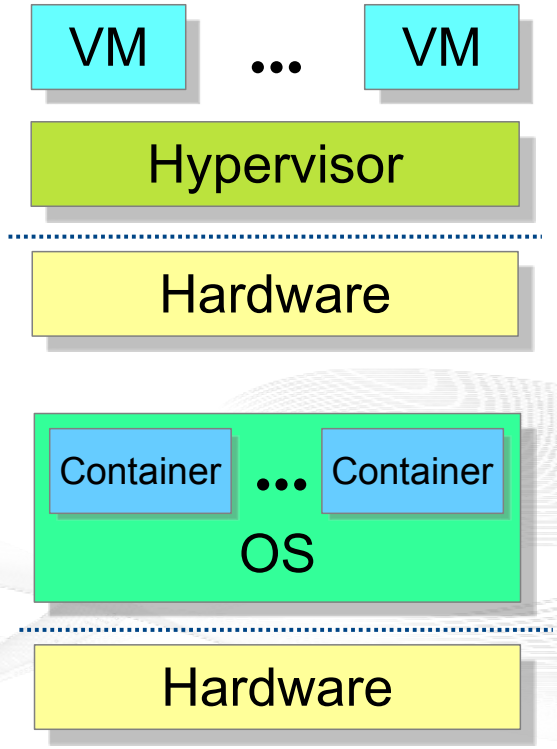
Hardware in a Cloud Data Centre

- User's local computer
  - Low utilisation
  - Main task: waiting for user input
- Idea: many computers, for a large number of users
  - **Computers in a data centre**
  - Usage by many users
  - Usage distribution over time
  - Scalability
  - High utilisation, **low costs**



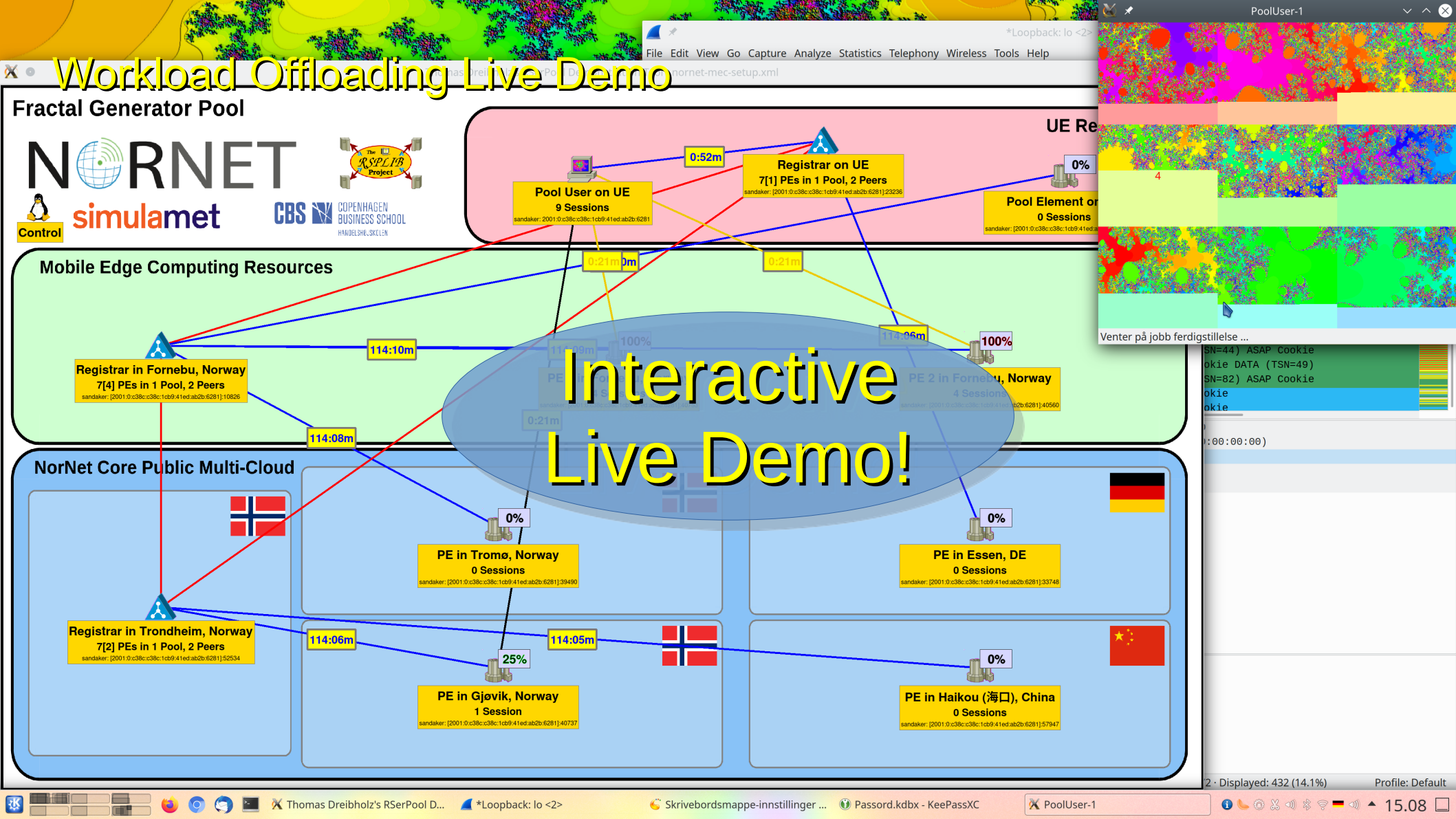
Virtual Machines and Containers

- Data centres can provide “virtual computers” for the users
- **Virtual Machines** (VM)
  - Full operating system (OS) on virtualised hardware
  - Own operating system and kernel version
  - Very flexible, but requires resources
- **Containers**
  - Containers run on shared OS kernel
  - Containers are shielded from each other (own view of processes, file system, networking, ...)
  - More lightweight than VM, but same OS/kernel



Latency and Fog/(Mobile) Edge Computing

- Cloud
  - Resources **somewhere**, where they are inexpensive
  - But network communication takes time → **latency**!
  - Speed limit: speed of light c≈ 3*10^8 m/s!
- Fog/(Mobile) Edge Computing (MEC, EC)
  - Adding resources nearby the user,
    e.g. computers at the user's Internet service provider
  - Backed by cloud resources
  - => **Low latency for latency-critical tasks**
- Offloading of work from user's system to edge/fog or cloud


Workload Offloading Live Demo
Fractal Generator Pool
NORNET
simulamet
Control
CBS COPENHAGEN BUSINESS SCHOOL
UE Re
Pool User on UE
9 Sessions
Registrar on UE
7[1] PEs in 1 Pool, 2 Peers
Pool Element or
0 Sessions
0:52m
0%
Mobile Edge Computing Resources
0:21m
114:10m
Registrar in Fornebu, Norway
7[4] PEs in 1 Pool, 2 Peers
100%
PE 2 in Fornebu, Norway
4 Sessions
114:08m
Interactive
Live Demo!
NorNet Core Public Multi-Cloud
PE in Tromø, Norway
0 Sessions
PE in Essen, DE
0 Sessions
Registrar in Trondheim, Norway
7[2] PEs in 1 Pool, 2 Peers
114:06m
114:05m
25%
PE in Gjøvik, Norway
1 Session
PE in Haikou (海口), China
0 Sessions
PoolUser-1
Venter på jobb ferdigstillelse ...
File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help
Profile: Default
15.08

What about Privacy?

- **Cloud** is quite convenient
  - **Inexpensive**, **scalable**, resources are available when needed
  - But what about privacy
- Cloud **data centres**
  - They are located somewhere, in a **country/region** with certain **regulations**
  - EU/EEA: General Data Protection Regulation (GDPR);
    USA: the US regulations; China: Chinese rules; / / / / …
- But what about the **network transfer** of data?
  - Over which countries/regions is data flowing?
  - Is this static, or does it change?
  - How can I find out details about my data flows?

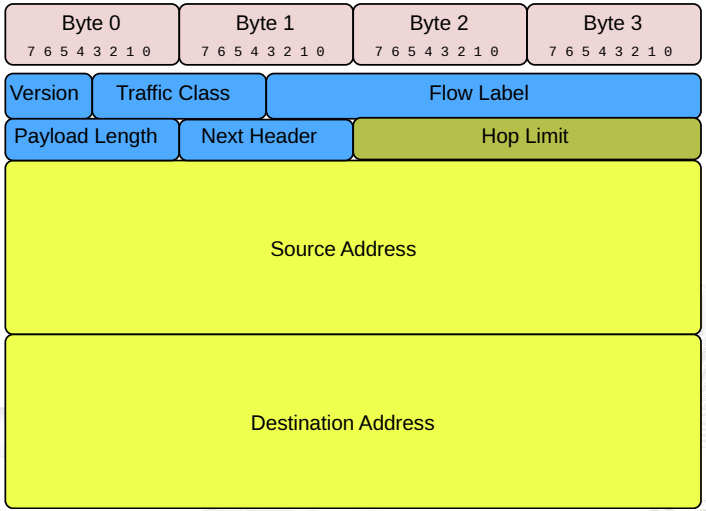
Network Communication – How does it work? (1)

- Analogon: sending items via post
  - **Pack** things into packets (with size limits)
  - **Add label** with recipient and sender
  - Take it to the local post office
  - **Each packet is routed individually**
  - Receiver picks up packets at his local post office
  - **Unpack** things from packets

Network Communication – How does it work? (2)

- Sending data via the Internet Protocol (IP)
  - **Encapsulate** (pack) data into packets
    - Size limit: Maximum Transmission Unit (MTU)
    - Usually ~1500 bytes
  - **Add header** including recipient (destination) and sender (source)
  - Send packets to your local network's router
  - **Each packet is routed individually**
  - Receiver gets packets from his local router
  - **Decapsulate** (unpack) data from packets

IP Header (version 6)



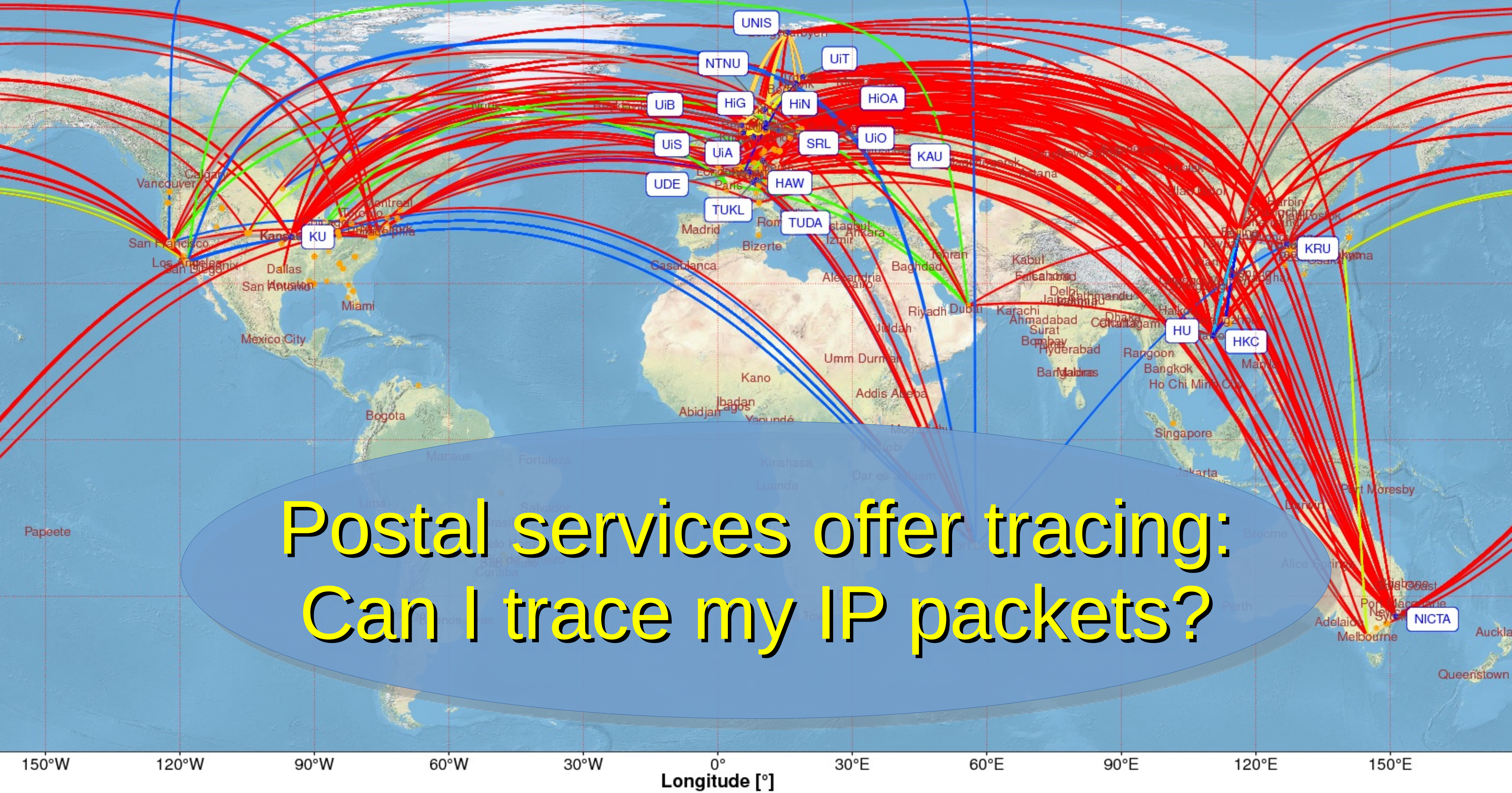
Addressing and Routing

- Postal addresses are hierarchical:
  - Country
  - Postcode and City
  - Street, Number
  - Name
- Routing: relevant recipient details
  - Int'l freight airline: only country/city
  - Domestic postal service: postcode
  - Postman: name, street and number

- Internet addresses are hierarchical:
  - Network ID
  - Host ID
- Routing: relevant receiver details
  - Trans-Atlantic cable provider:
    only aggregated network IDs of ISPs
  - Local ISP:
    aggregated network ID of customers
  - Local router:
    knows networks and devices at home

A computer networks course is recommended for more details!

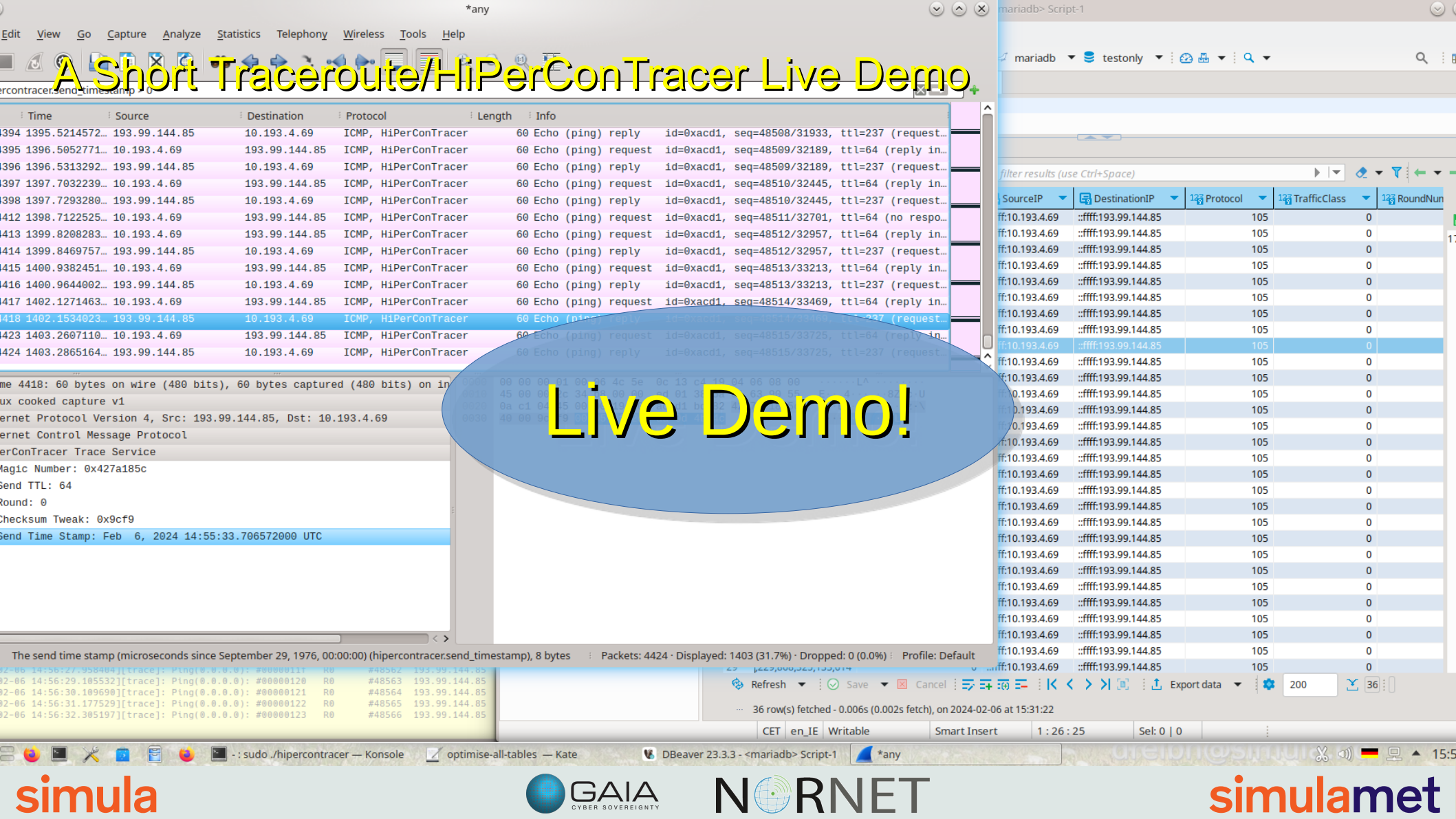
Postal services offer tracing: Can I trace my IP packets?



Traceroute and HiPerConTracer

- Traceroute
  - Send packets from source to destination (as usual)
  - Limit the number of intermediate stations (routers)
  - If destination is not reachable within the limit, the last router sends error
  - Result: sequence of all routers' addresses + known source/destination
  - Note: IP address ≠ geo-location!
- Larger-scale Traceroute runs
  - HiPerConTracer framework
  - See https://www.nntb.no/~dreibh/hipercontracer/

A Short Traceroute/HiPerConTracer Live Demo



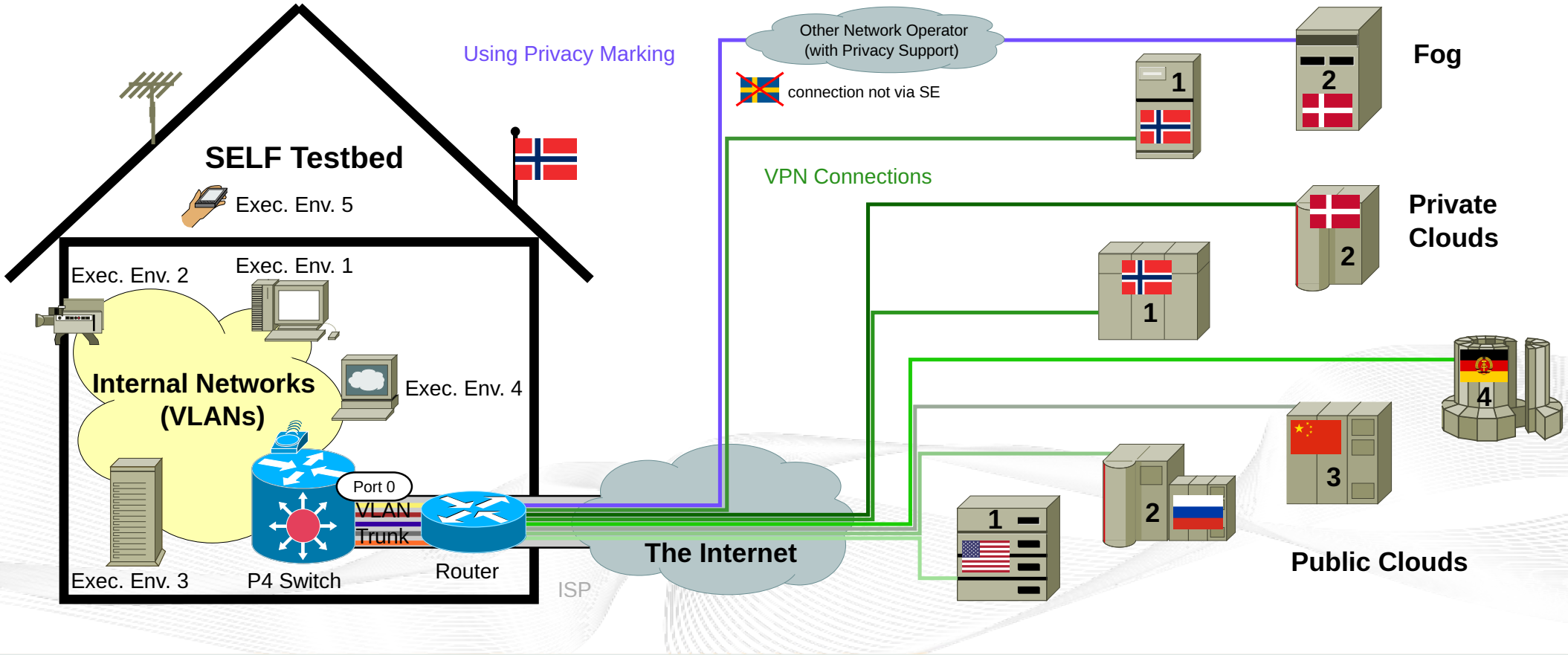
Live Demo!

simula GAIA CYBER SOVEREIGNTY NORNET simulamet

What can be done with this?

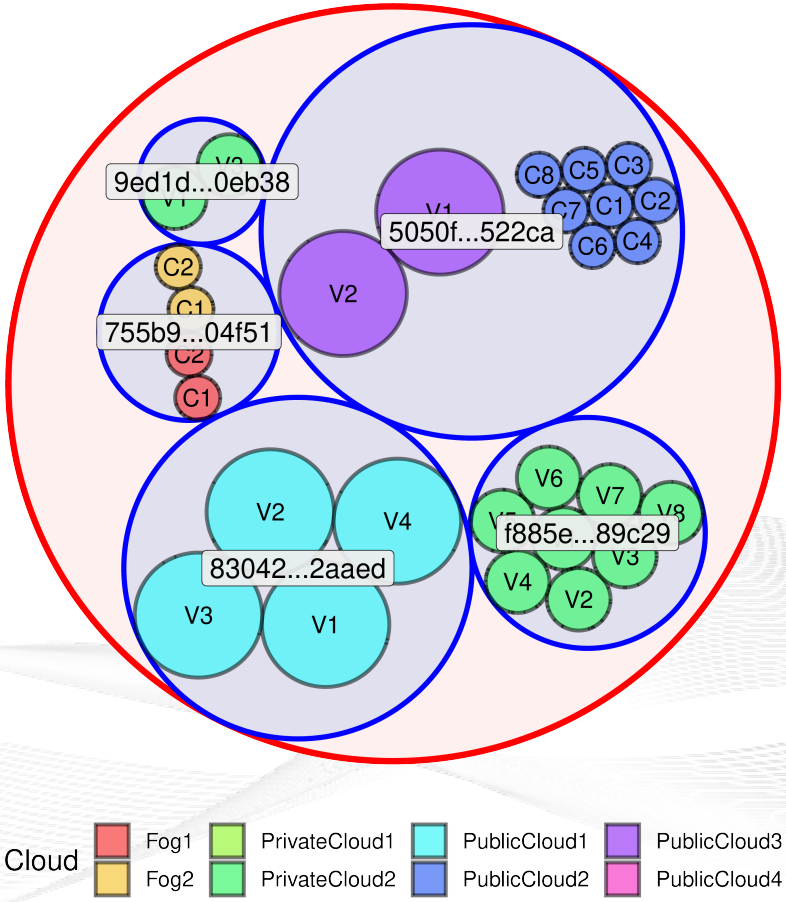
- Continuous, larger-scale measurements
  - See changes over time (long-term, short-term)
  - Perform geo-location of addresses
    - Lookup in databases
    - Triangulation measurements from known vantage points
- Idea: Secure Embedded Living Framework (SELF)
  - Give user some control over desirable/undesirable routes
  - Restrict connectivity of devices/groups of devices

Secure Embedded Living Framework (SELF)



Dynamic Execution Environments

- Execution Environments (EE):
  - Assign devices, access to clouds/fogs, etc.
  - Created/removed on demand
  - Dynamic scaling (more/less resources)
- Example:
  - EE1: security cameras, processing in Fog1 + Fog2
  - EE2: some AI application, compute resources in PublicCloud1
  - ...
- Different EEs do not interact
  - Security/privacy issue in one EE does not affect other EEs



Conclusions

- Privacy is an important and broad topic!
  - Cloud/fog resources located somewhere
  - Network communication is very interesting as well
    - Routing changes over time
    - Routing may take quite unexpected detours
      (via different countries, regions, network providers, ...)
  - Some ideas for improvements → SELF concept
- Opportunities for Bachelor/Master topics
  - Collaboration between SimulaMet and CBS

simulamet

CBS COPENHAGEN BUSINESS SCHOOL HANDELSHØJSKOLEN

Literature

Mazumdar, S. and Dreibholz, T.: "Towards A Data Privacy-Aware Execution Zone Creation on Cloud/Fog Platform", in Proceedings of the 49th Euromicro Conference Series on Software Engineering and Advanced Applications (SEAA), pp. 140–149, Durrës/Albania, September 2023, URL: https://web-backend.simula.no/sites/default/files/2023-10/SEAA2023.pdf.

Dreibholz, T. and Mazumdar, S.: "Towards a Lightweight Task Scheduling Framework for Cloud and Edge Platform", in Internet of Things, vol. 21, Elsevier, April 2023, URL: https://web.archive.org/web/20230517075030/https://www.simula.no/file/iot2023pdf/download.

Mazumdar, S. and Dreibholz, T.: "Towards a Privacy Preserving Data Flow Control via Packet Header Marking", in Proceedings of the 24th IEEE International Conference on High Performance Computing, Data, and Analytics (HPCC), pp. 1509–1516, Chengdu, Sichuan/People's Republic of China, December 2022, URL: https://web.archive.org/web/20230520172441/https://www.simula.no/file/hpcc2022pdf/download.

Mazumdar, S. and Dreibholz, T.: "Secure Embedded Living: Towards a Self-contained User Data Preserving Framework", in IEEE Communications Magazine, vol. 60, pp. 74–80, November 2022, URL: https://web.archive.org/web/20230920185748/https://www.simula.no/file/commmag2022pdf/download.

Dreibholz, T. and Mazumdar, S.: "Find Out: How Do Your Data Packets Travel?", in Proceedings of the 18th IEEE International Conference on Network and Service Management (CNSM), pp. 359–363, Thessaloniki, Greece, November 2022, URL: https://web.archive.org/web/20230920185748/https://www.simula.no/file/cnsm2022pdf/download.

Dreibholz, T. and Mazumdar, S.: "A Demo of Workload Offloading in Mobile Edge Computing Using the Reliable Server Pooling Framework", in Proceedings of the 46th IEEE Conference on Local Computer Networks (LCN), Demo presentation, Edmonton, Alberta/Canada, October 2021, URL: https://web.archive.org/web/20230920185748/https://www.simula.no/file/lcn2021-rserpool-webpdf/download.


Thank you for your attention!
Any questions?
Thomas Dreibholz
dreibh@simula.no
https://www.simula.no/people/dreibh