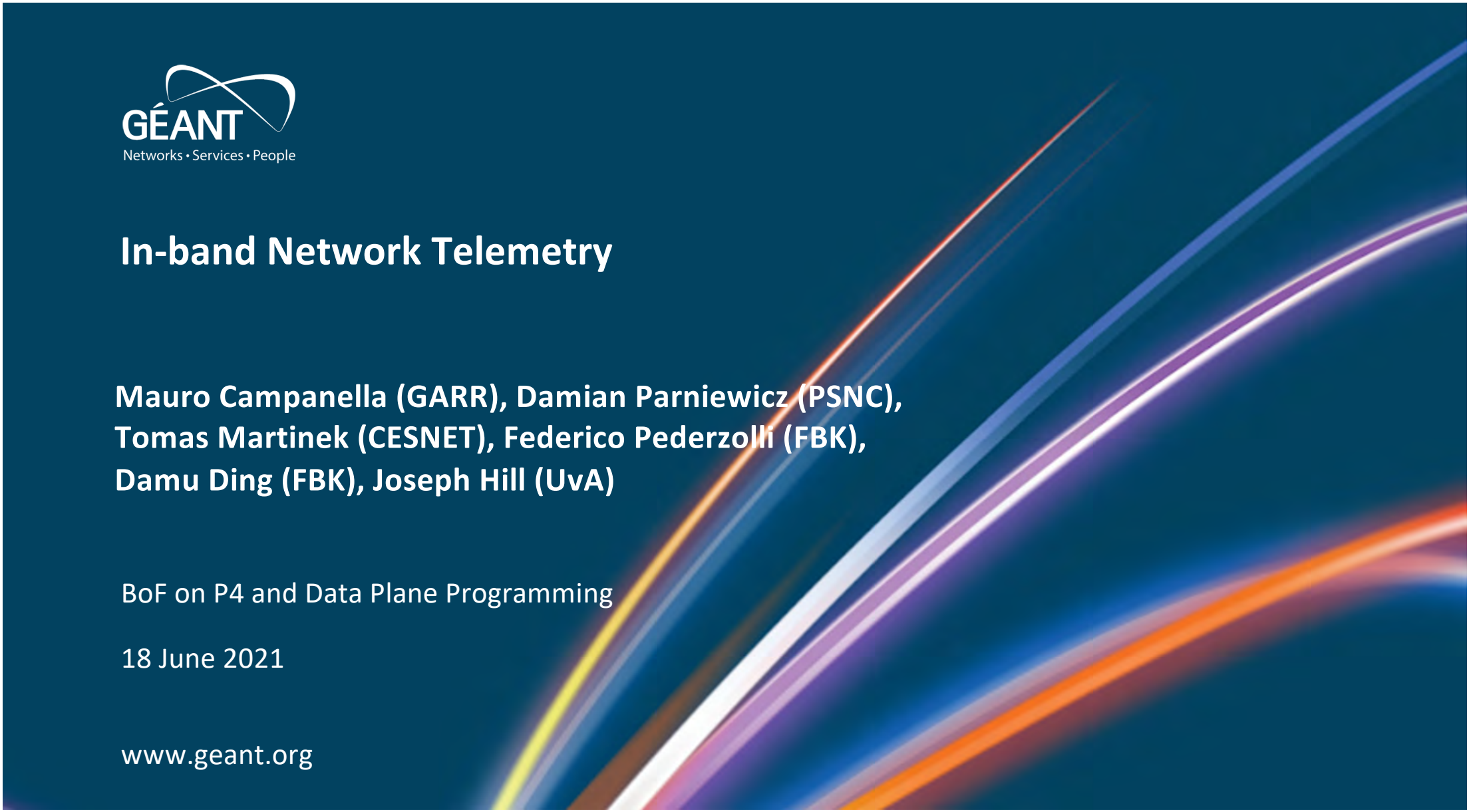


In-band Network Telemetry

Mauro Campanella (GARR), Damian Parniewicz (PSNC), Tomas Martinek (CESNET), Federico Pederzolli (FBK), Damu Ding (FBK), Joseph Hill (UvA)

BoF on P4 and Data Plane Programming

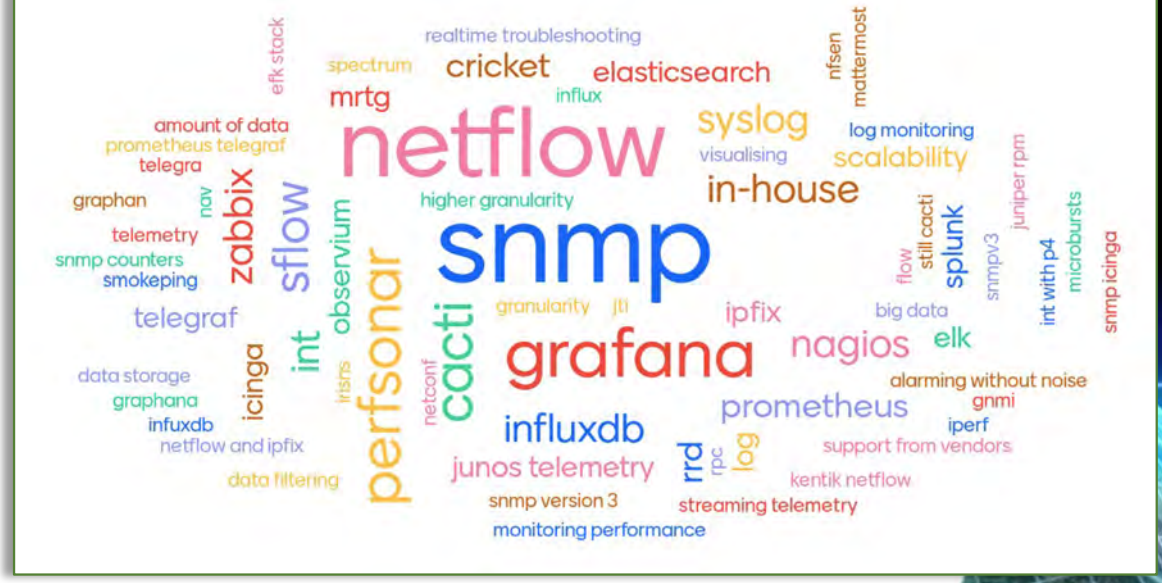
18 June 2021

www.geant.org

Why INT ?

- Simple, effective, application of data plane programming to network monitoring
- INT extends 'classical' monitoring and streaming telemetry to improve granularity in time, space and subject (choice of flow / packet / protocol…) and allows local **data manipulation,** realtime

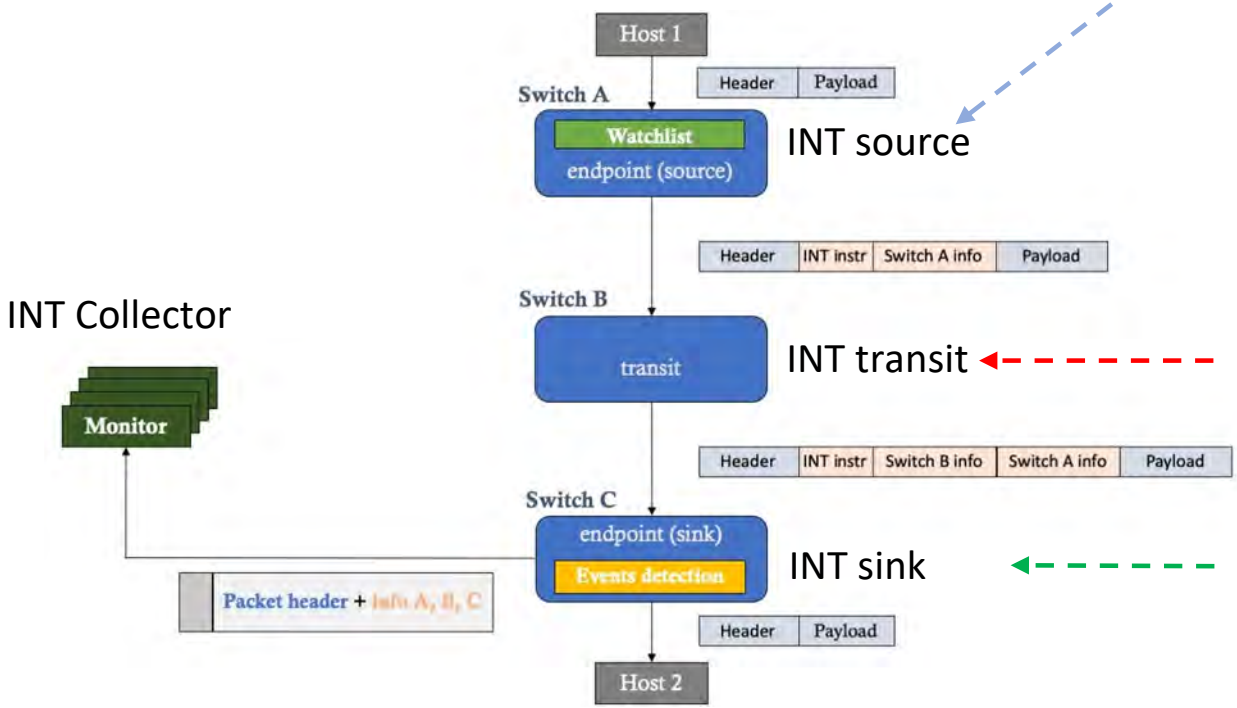
From participants to the Telemetry and Big Data Workshop: **monitoring protocols most used** today



- Explore network behavior at finer (µs, ns) time graularity (new **knowledge** for debugging and controlling Terabit, clouds)

In-Band Network Telemetry (INT) summary

INT was specified by the P4 language community to provide very detailed information on network behaviour

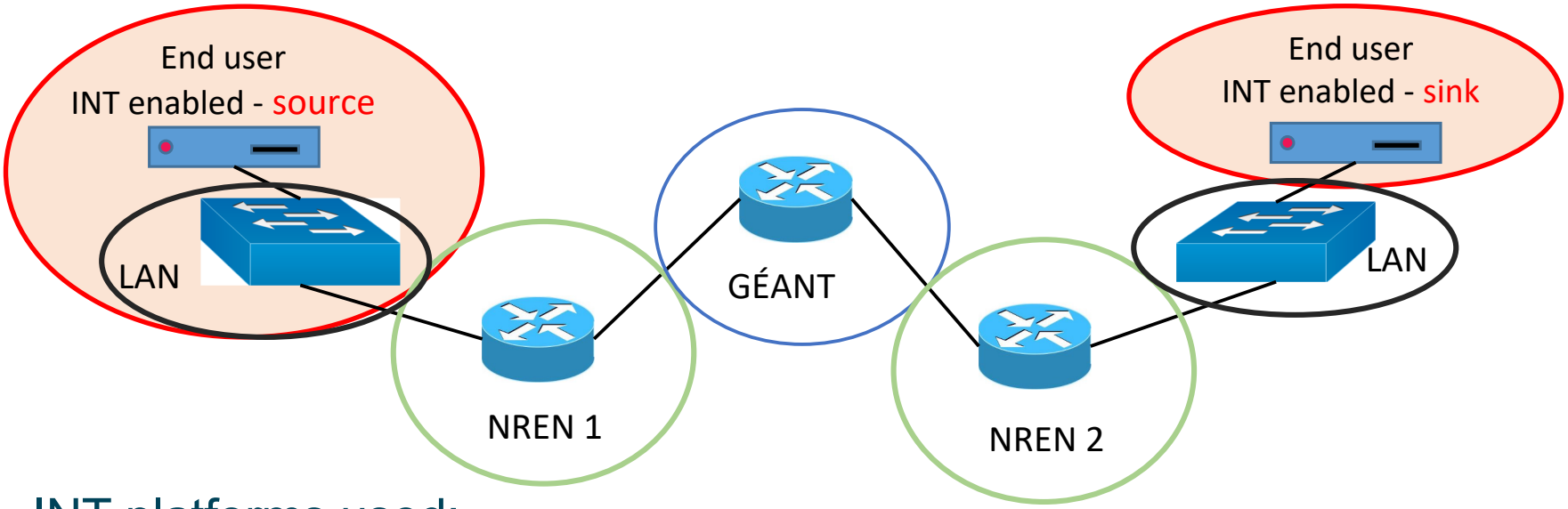


INT **source node** adds a small **INT header** (~50B) in **every chosen packet** with local information (Interfaces and switch IDs, Timestamps, Link and queue utilization e.g.)

INT transit nodes may add, local info, manipulate headers

The **INT sink** node exports INT data to the collector for storage/visualization

We acted as end-users to measures own traffic IPDV, Loss,…



INT platforms used:

Edgecore Wedge100BF-32X Arista 7170-32c

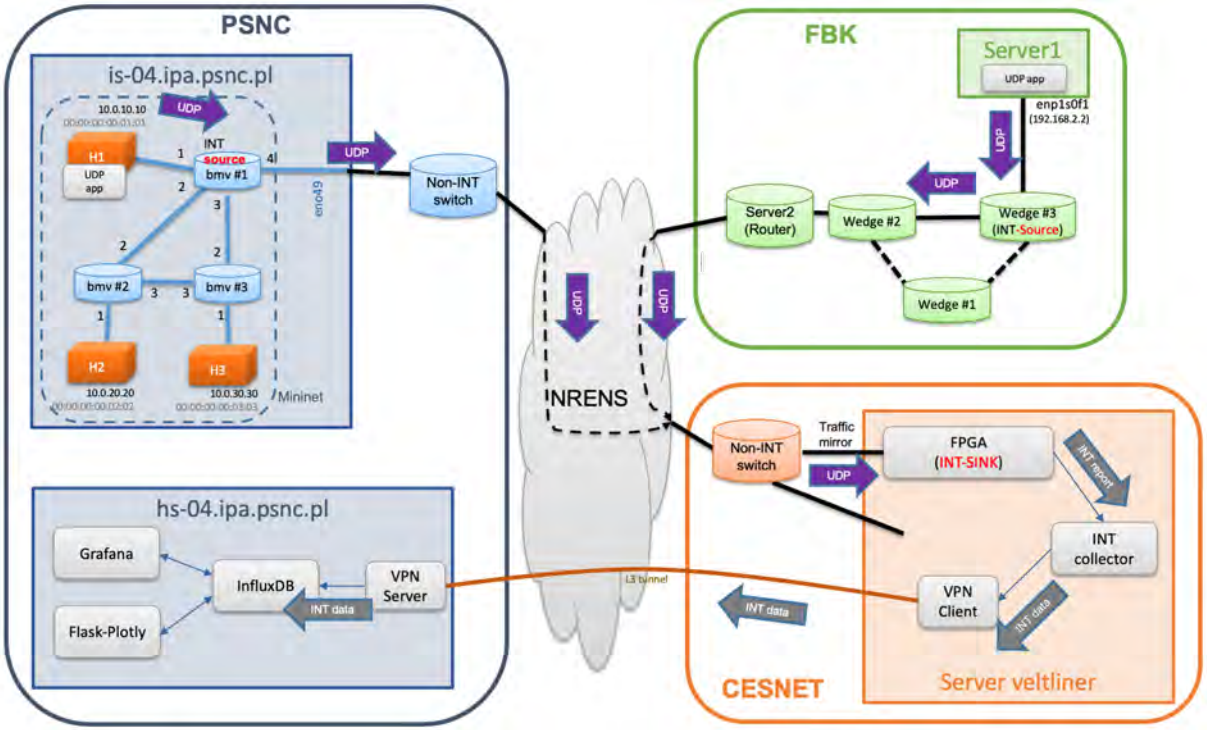
Tofino (Barefoot/INTEL)

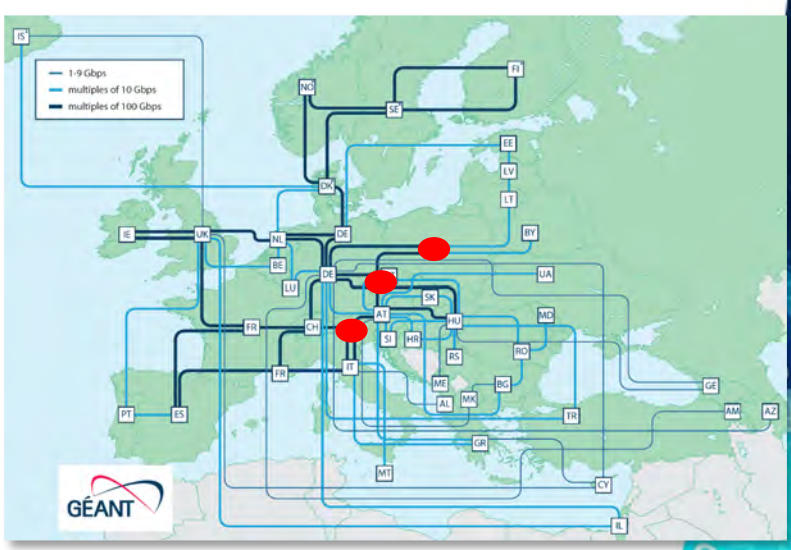
BMv2 virtual
P4 software switch

P4 on DPDK
implementation

FPGA

INT: testbed over production NREN networks





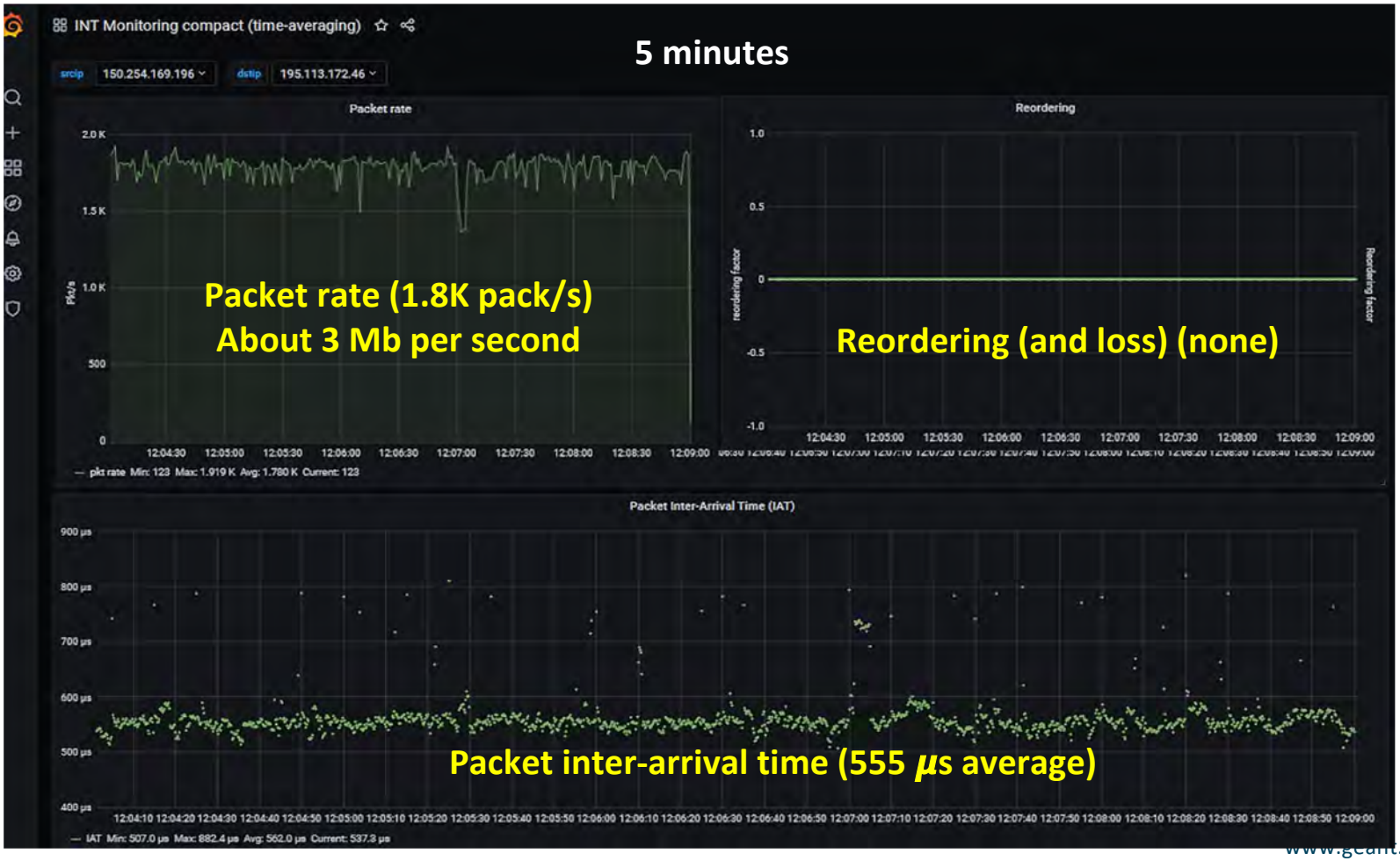
- 3 switch types
- UDP packets flow on NRENs networks
- Collected INT data in CESNET is sent back to PSNC for collection and presentation.

INT Platforms: Lessons Learned

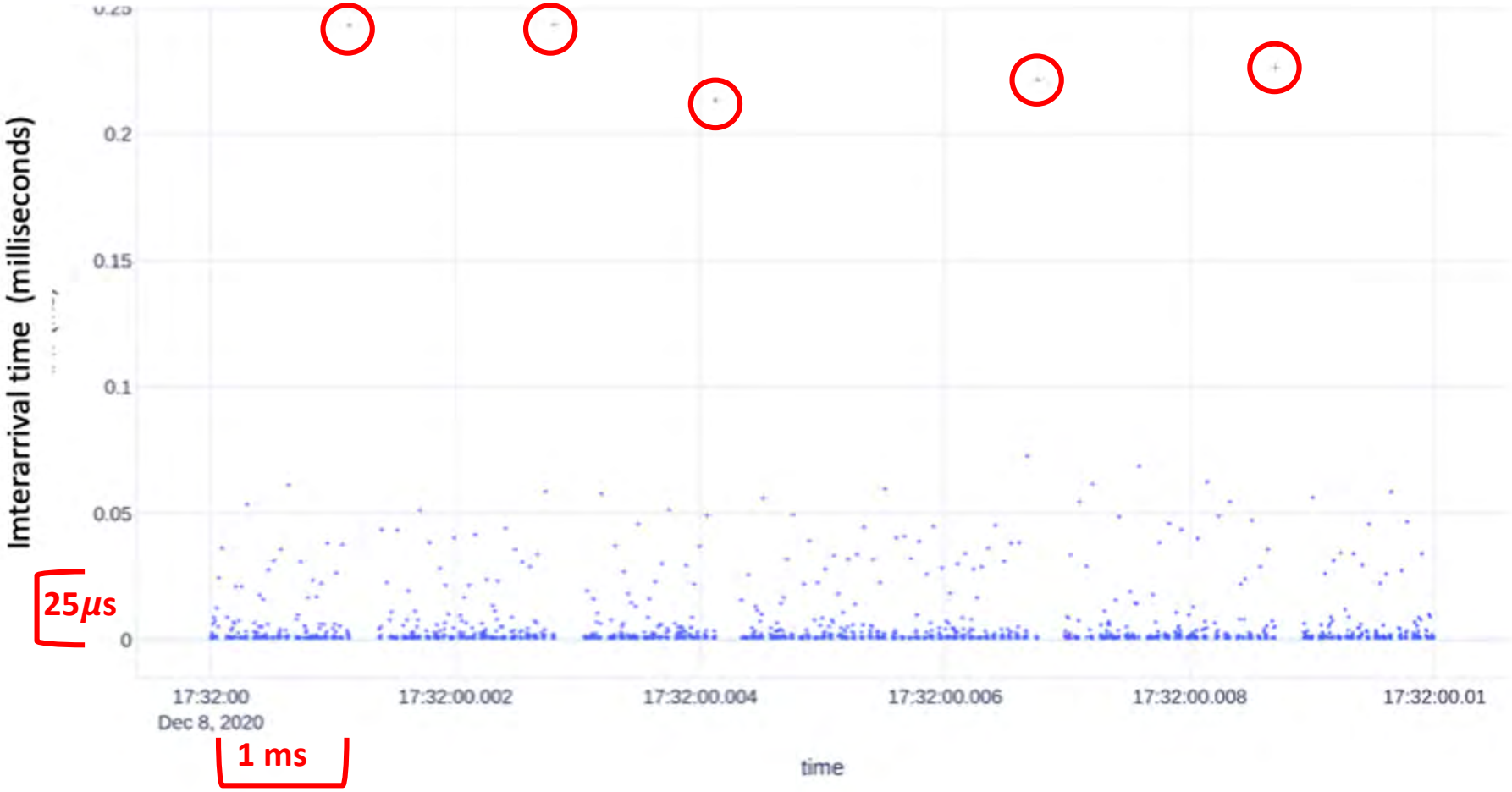
Pick the solution that works for the use case

- Bmv2/mininet: good to start initial P4 code development
  - Performance ceiling, virtual routing
- Tofino switches: potentially feature-rich for P4
  - Clock synchronisation, licensing, complexity issues
- FPGA card: fast, flexible HW
  - P4 compiler vital; CESNET compiler for P4_14
  - HW expertise may be required for some features
- INT-DPDK:
  - Promising performance up to ~10G
  - Needs careful selection of NICs

5 minutes of the INT monitored flow from PSNC to CESNET



10 ms of inter-arrival packet time, 1 UDP flow of 260 K pps (~3.3 μs average) FBK to CESNET (1 packet is ~5 μs – 1 Gb LAN)



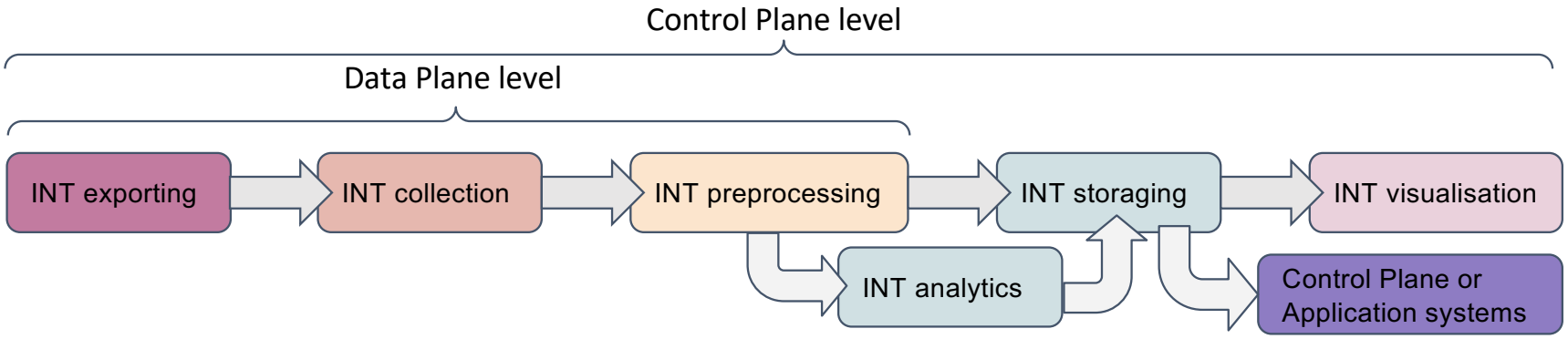
Data “velocity” in In-Band Telemetry

Assuming that every flow packet is monitored:

Flow rate	Only 64B packets (+20B interpacket gap)	Only 1518B packets (+20B interpacket gap)	Only 9018B packets (+20B interpacket gap)
100Mbps	149K reports/s ✓	8.13K reports/s ✓	1.38K reports/s ✓
1Gbps	1.49M reports/s	81.3K reports/s ✓	13.8K reports/s ✓
10Gbps	14.9M reports/s	813K reports/s	138K reports/s ✓
100Gbps	149M reports/s	8.13M reports/s	1.38M reports/s

- Performance must be scaled by the number of simultaneously monitored connections
- INT reports require near real-time data processing or batch processing:
  - Generate events (anomalies)
  - Calculate aggregated statistics
  - Provide visualisation

Improving INT scalability for high-rate flows, mutiple flows, more data)



P4 language

eXpress Data Path

Data Plane Development Kit

extended Berkeley Packet Filter

Remote Direct Memory Access

Data filtering

Key events finding

Data aggregation

Time resolution

Kafka Streaming cluster

ElasticSearch cluster

InfluxDB Enterprise cluster

Spark cluster

Data Plane level

More generality and flexibility, bigger server clusters

Many limitations, hardware special requirements

Control Plane level

INT activity status and next step

- INT **P4 code,** INT Spec 1.0, is **available from Github** for Tofino, BMv2, FPGA tests, and DPDK soon
- Data collection and presentation tools and configuration available

Next steps:

- Planning improvement of **clock synchronization**
- Development of a tool based on virtual nodes with a complete INT system
- Systematic measurements and analysis of behavior of production networks

Summary

- **INT** (and Data Plane Programming) (using P4) is not business-as-usual, requires specific expertise, however it offers a innovative technology for **monitoring, debugging and providing information to control plane, in real time.**
- P4/INT is more and more available in various platforms (switches and linecards, software)
- INT is a powerful **magnifying glass** on network behaviour and can operate e2e between cooperating domains without imposing complex agreements for the control plane
- Time synchronization between nodes is important
- As a function of the use case, the INT/P4 use may generate and require handling of **large amount of "raw" data**, to be used for analytics and more. It implies the development of further insight, **knowledge** and specific tools and equipment to scale.

More information

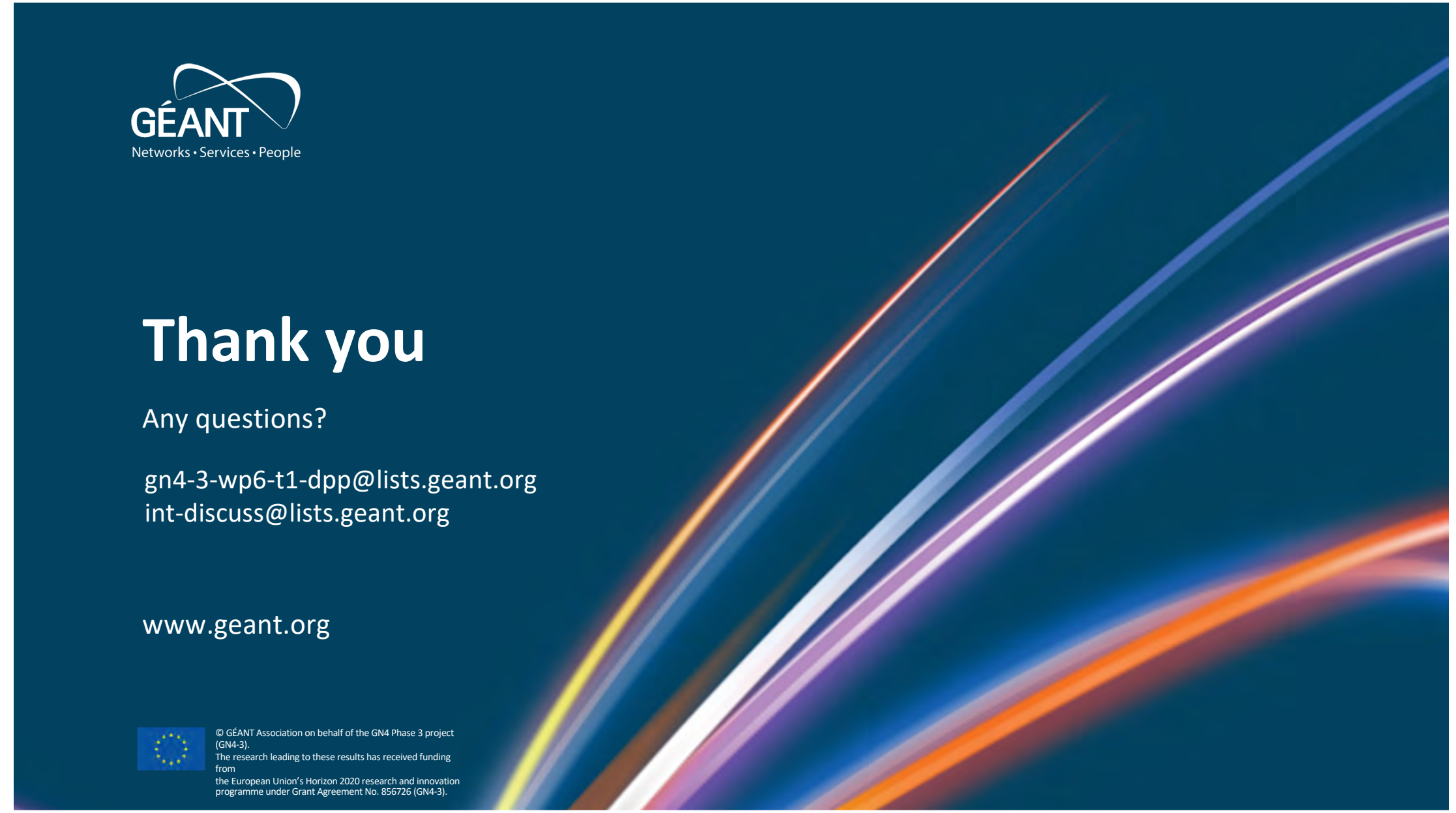
- **Data Plane Programming / INT GEANT web page**
  https://wiki.geant.org/display/NETDEV/INT
  Includes all documents produced and a pointer to GitHub INT P4 code

- **Mailing list:** https://lists.geant.org/sympa/subscribe/int-discuss,

- **White Paper INT Tests in NREN networks** – DPP WP6 T1 white paper
  https://www.geant.org/Resources/Documents/GN4-3_White-Paper_In-Band-Network-Telemetry.pdf

- The GÉANT **First Telemetry and Big Data Workshop**
  https://wiki.geant.org/display/PUB/Telemetry+and+Big+Data+Workshop

- **Paper**: "In-Network Volumetric DDoS Victim Identification Using Programmable Commodity Switches", F. Pederzolli, M. Campanella and D. Siracusa, in IEEE Transactions on Network and Service Management, Vol. 18, Issue: 2, June 2021, page: 1191-1202, DOI: 10.1109/TNSM.2021.3073597 and at https://arxiv.org/abs/2104.06277



The research leading to these results has received funding from
the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 856726 (GN4-3).