

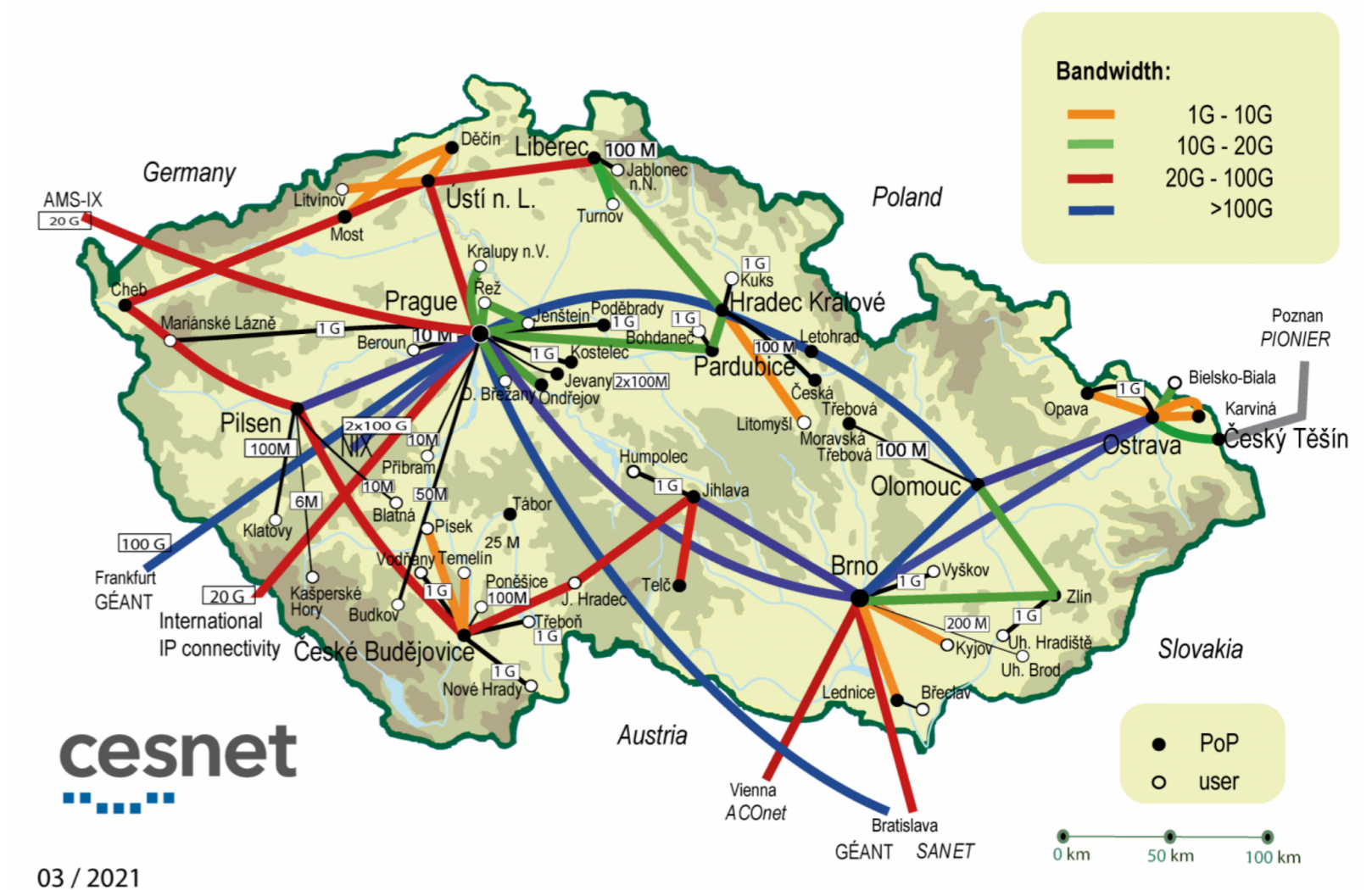
Introduction and Motivation

Machine learning methods are often used for monitoring and analysing network traffic. However, they require **high quality datasets**, which can be problematic to create. Traffic Capture Infrastructure (TCI) is a system for the distributed creation and automatic processing of such datasets. Users can:

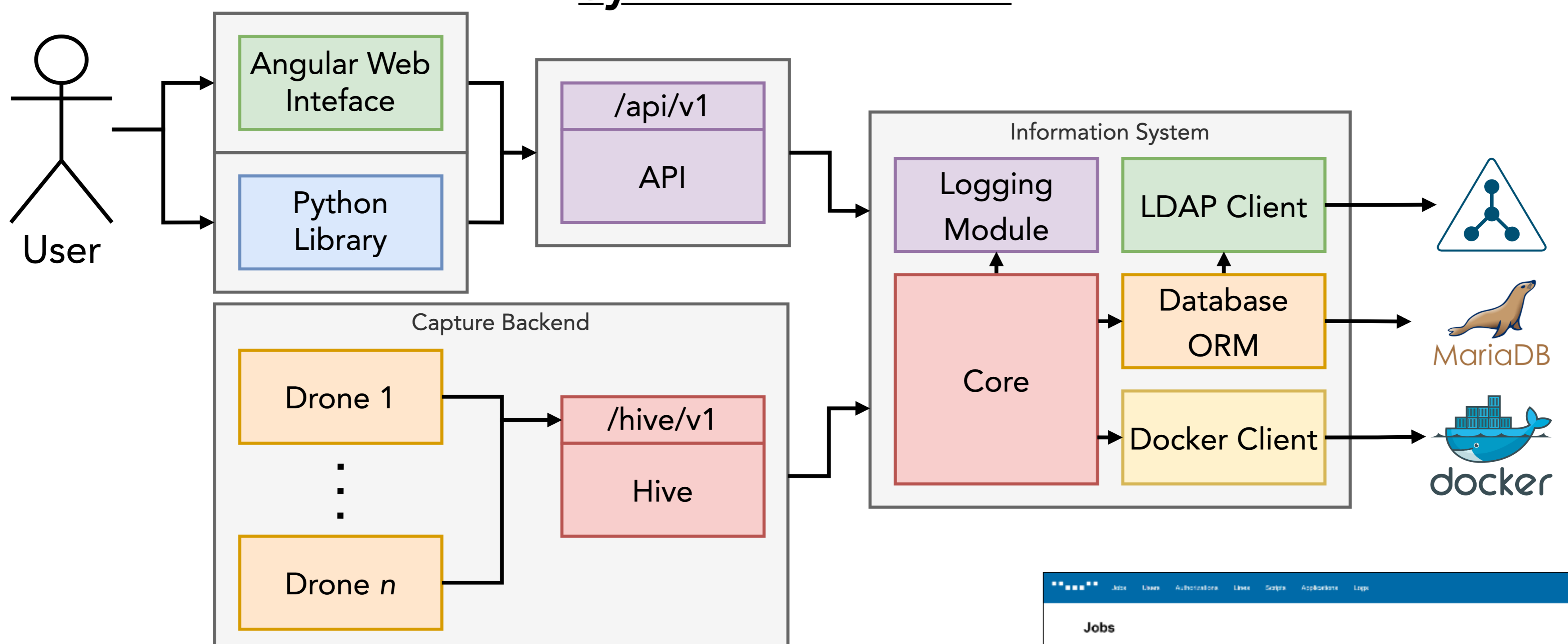
- Create **capture jobs on high-throughput network lines** (support for up-to 400Gbit network throughput using NDP).
- Automatically **process the data** using an air-gapped Docker container.
- Check the status of their capture jobs using the **web interface**.
- Automatically start capture jobs using the **Python client library**.
- Support for deploying on **Cisco Catalyst** devices.

Existing Publications and Deployments

The TCI system is deployed on the backbone infrastructure of the CESNET2 network, and datasets created using this system served as the basis for **11 papers and publications**, with further papers being prepared. TCI is also used by the **CESNET-CERTS** security team for network incident analysis.



System Architecture



1. The **User** uses the **Angular Web Interface** or the **Python Client** library for creating a new capture job on selected capture points.
2. The **Angular Web Interface** sends this information to the **REST API** which verifies this request, and passes it on to the **Information System**.
3. The **Information System** logs this information, and creates the new capture job, which is sent to the **Capture Backend** using a **SocketIO** connection.
4. Once all data is captured, the **Information System** collects the data and processes it using the **Docker** client to ensure privacy and user convenience.
5. Once the data is processed, the user can download it.

Acknowledgements

The TCI project was developed as part of the project "FETA - Flow-based Encrypted Traffic Analysis" (VJ02010024), which was supported by the Ministry of the Interior of the Czech Republic.

