Real-time Flight Data Processing with timely-dataflow

R244 — Mini Project

Andreas Pletschko (ap2535@cam.ac.uk) — 03.12.2025

Project Motivation

Background — Flight ADS-B Data

- Airplanes are required to periodically broadcast their ADS-B information (positions, …)
- Certain platforms collect these broadcasts and make them available online, such as the **OpenSky Network** [1]
  - Historical database: Various 5-day chunks of ADS-B data feed between 2023 and 2024
  - Live API: ~60,000 ADS-B transmissions per minute
- Various analytics:
  - Airspace density/Congestion analysis
  - Collision warnings/detection, …

Project Goal

Large amounts of streamed data + iterative computations make **timely-dataflow** attractive, compared :

- Implement a pipeline to process ADS-B data for **congestion analysis** in **timely-dataflow**, measuring metrics such as throughput, latency, CPU/memory utilization
- Extensions
    - Possible collision detection (using distributed spatial indices, such grid-based partitions)
    - Comparison with other frameworks, such as Spark (batching vs streaming) or Flink (JVM vs native Rust)

Project Plan

- **Day 1:** Set up timely-dataflow, get comfortable with the programming model
- **Day 2:** Set up ADS-B ingestion pipeline, get comfortable with the data and the API
- **Day 3:** Implement congestion analysis pipeline
- **Day 4-5:** Implement one or more of the possible extensions (collision detection, comparison with Flink/Spark)
- **Days 6:** Benchmark and analyze results
- **Day 7:** Write report

Questions?

References

[1] Matthias Schäfer, Martin Strohmeier, Vincent Lenders, Ivan Martinovic, Matthias Wilhelm. „Bringing Up OpenSky: A Large-scale ADS-B Sensor Network for Research“. In Proceedings of the 13th IEEE/ACM International Symposium on Information Processing in Sensor Networks (IPSN), pages 83-94, April 2014.