Acquisition Protocol - Impact on Real-time Data Acquisition System Performance

nanometrics

Andrew Owen Moores* (andrewmoores@nanometrics.ca), Michael Laporte*, Ben Tatham*, Michael Perlin*, Francis Tong*, Drew Gibson* (* Nanometrics Inc.)

SUMMARY

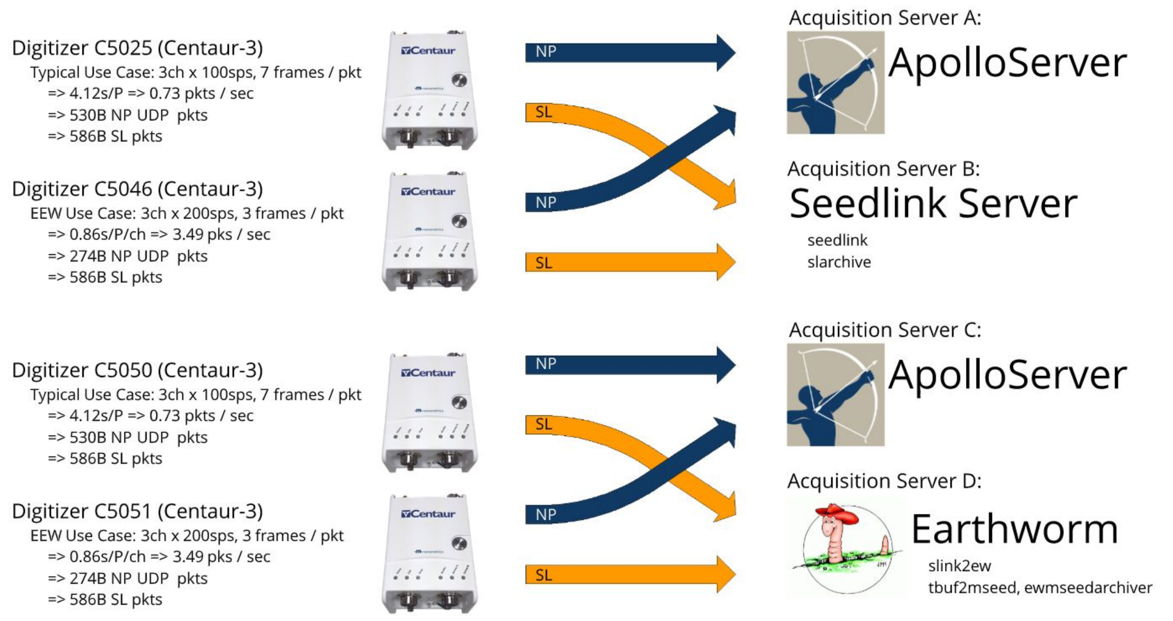
A fundamental element of real-time mission critical seismic monitoring networks is the data acquisition system, comprising the underlying protocol and the telemetry solution. Selection of the acquisition protocol can have significant impact on key network performance metrics, as well as operational cost and even station and data center design.

The performance of the Nanometrics NP UDP and SeedLink acquisition protocols is assessed in various adverse conditions by examining bandwidth utilization, data latency and acquisition robustness (data completeness).

In addition, protocol functionality and features, including support for multiple data types and state-of-health, are assessed for system impact on options for station, telemetry, and data center design as well as the overall functionality of the system solution.

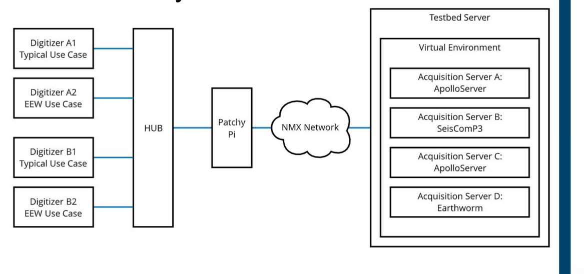
TEST SETUP

- Utilized the Nanometrics product development long-term integration testbed
- Data loggers: 4 x Centaur-3 Digitizers
  - 3 channels, shorted inputs, supports both NP UDP and Seedlink acquisition, 100kbps throttle
- Telemetry & Networking
  - Ethernet connections with inline custom "Patchy Pi" network device
  - Patchy Pi simulates a poor communications link by periodically disabling network connectivity
  - Identical physical routing between each data source and destination
- Acquisition Systems: Four servers operating independently, implemented as virtual machines on the same physical server
  - Acquisition A: ApolloServer
    - Acquires data from Centaurs C5025 and C5046 via NP UDP
  - Acquisition B: SeedLink Server (`seedlink`, `slarchive`)
    - Acquires data from Centaurs C5025 and C5046 via Seedlink
  - Acquisition C: ApolloServer
    - Acquires data from Centaurs C5050 and C5051 via NP UDP
  - Acquisition D: Earthworm (`slink2ew`, `tbuf2mseed`, `mseedarchiver`)
    - Acquires data from Centaurs C5050 and C5051 via Seedlink streaming



- Performance Monitoring
  - Bandwidth: derived from pcap files captured on acquisition servers
  - Latency (*normalized to average per-sample latency):
    - ApolloServer SOH reporting
    - SeedLink Server: slarchive metrics*
    - Earthworm: sniffwave metrics*
  - Data Completeness: post-experiment examination of miniseed archives

Physical Connections



EXPERIMENT 1: Normal Operation

Under stable acquisition conditions:

- NP uses less total bandwidth - significantly less for the EEW use case which requires less than half that of SeedLink. This is expected since the overhead associated with UDP is less than that for TCP and, for the EEW case, SeedLink requires padding of the smaller packets to 512B.
- Both NP systems reported comparable latency, while the SeedLink system was slightly higher and the EW higher still. The reason for the different results is under investigation.
- All systems have 100% data completeness. As expected, minor error correction activity is visible in the NP-acquired data sets, due to the data being out of order. In a Production environment, the archiving system would be configured to order the data.

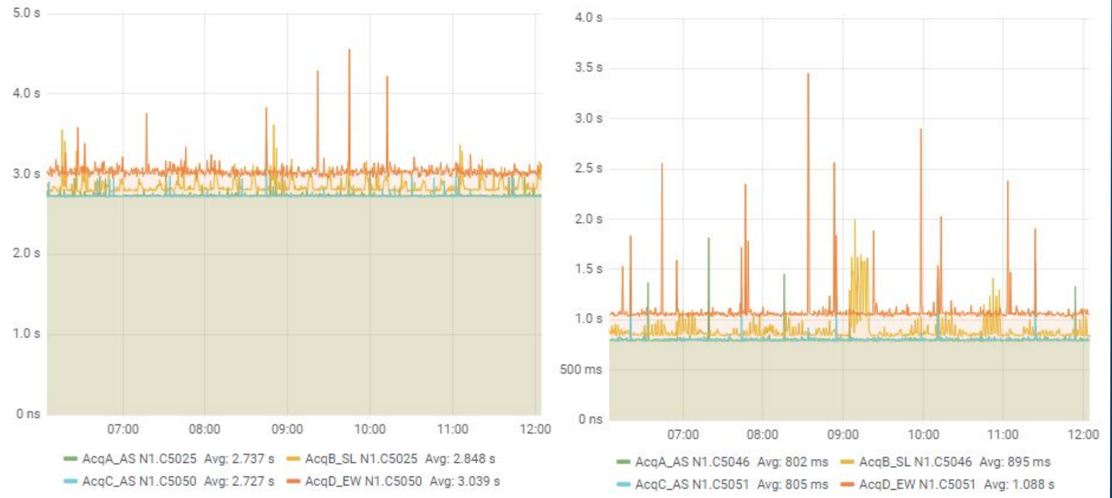
Bandwidth

Total bandwidth utilization estimates (combined inbound and outbound) were calculated by measuring the total data transferred over a fixed period of stable acquisition.

Typical Use Case Total Bandwidth Baseline:
- AcqA: 3472 bps, AcqB: 3832 bps
- AcqC: 3465 bps, AcqD: 3798 bps

EEW Use Case Total Bandwidth Baseline:
- AcqA: 8029 bps, AcqB: 18227 bps
- AcqC: 8044 bps, AcqD: 18195 bps

Latency



Typical Use Case Latency Baseline

EEW Use Case Latency Baseline

Data Completeness

Data completeness plots generated using `obspy-scan`

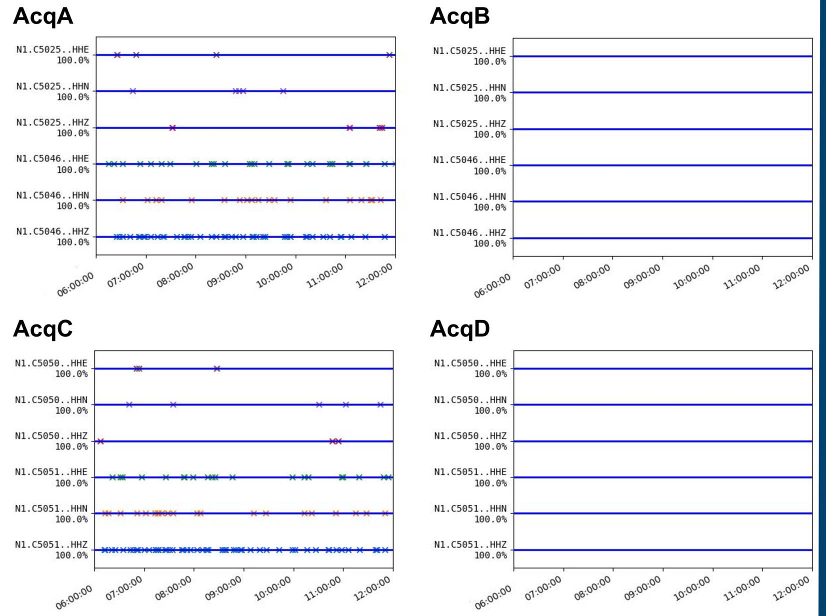
Crosses: Start of continuous data segments (can indicate out of order data due to gap recovery)

Horizontal Lines: Data

Vertical Red Lines: Gaps

Vertical Blue Lines: Overlaps

Crosses in NP-acquired plots (AcqA and AcqC) indicate gap recovery activity.

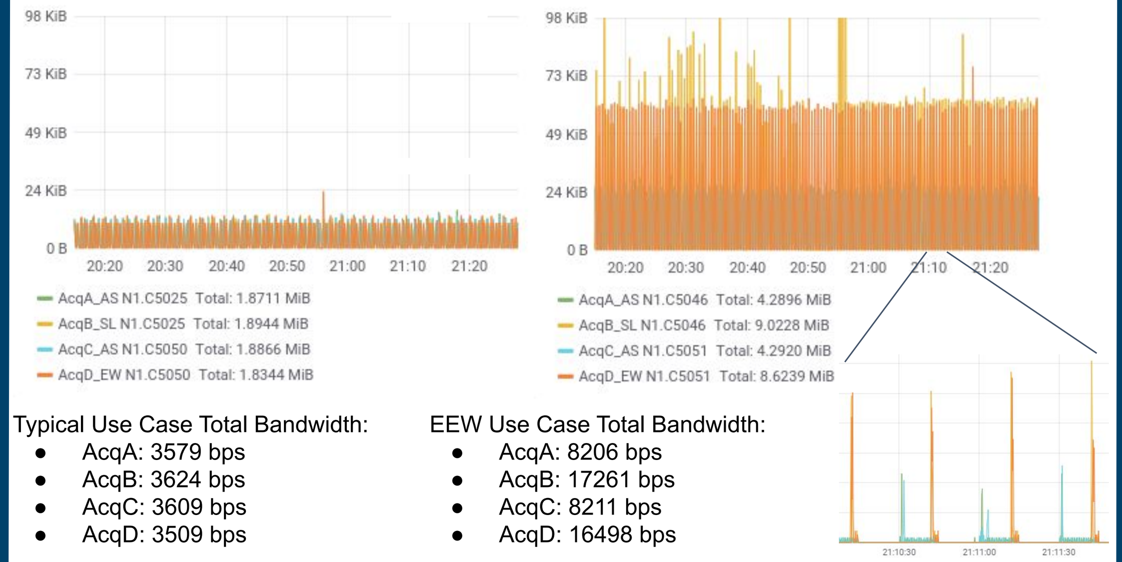


EXPERIMENT 2: 15s Duty Cycle

From stable acquisition, connectivity between the digitizers and acquisition systems was cycled, 15s ON / 15s OFF, for a period > 1 hour until it was permanently disconnected and performance assessed.

- Bandwidth utilization is generally consistent with baselines, with some variation. This is expected due to gap recovery activity (increased traffic) and missed data (decreased traffic).
- Latency varies significantly within the dataset, particularly for the SeedLink based systems. This is not unexpected given the TCP reconnect cycle must align with when the link is up.
- The NP-acquired systems had scattered gaps on half the channels. The SL system had all data except for the end of the period. The EW system had a few scattered gaps on all channels.

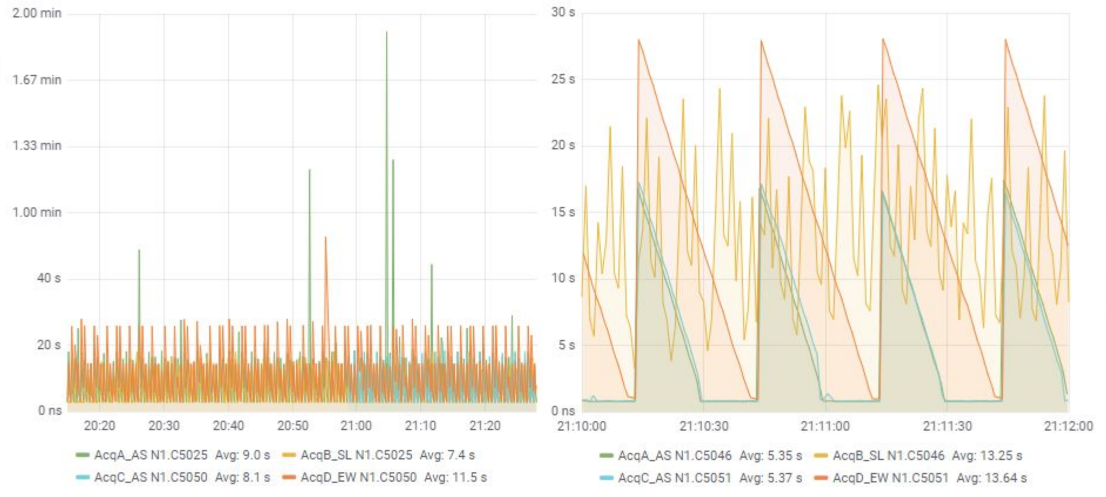
Bandwidth



Typical Use Case Total Bandwidth:
- AcqA: 3579 bps
- AcqB: 3624 bps
- AcqC: 3609 bps
- AcqD: 3509 bps

EEW Use Case Total Bandwidth:
- AcqA: 8206 bps
- AcqB: 17261 bps
- AcqC: 8211 bps
- AcqD: 16498 bps

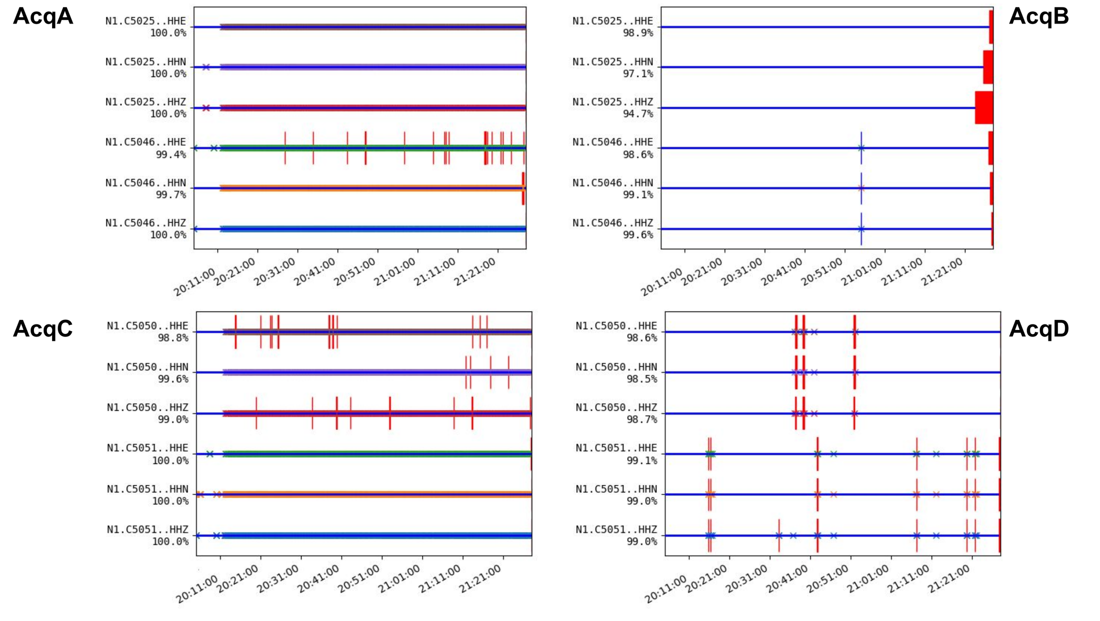
Latency



Typical Use Case Average Latency

EEW Use Case Average Latency (zoomed)

Data Completeness



EXPERIMENT 3: Large Outage Recovery

From stable acquisition, connectivity between the digitizers and acquisition systems was interrupted for a period of 12 hours. The connection was restored and the systems allowed to return to steady state operation.

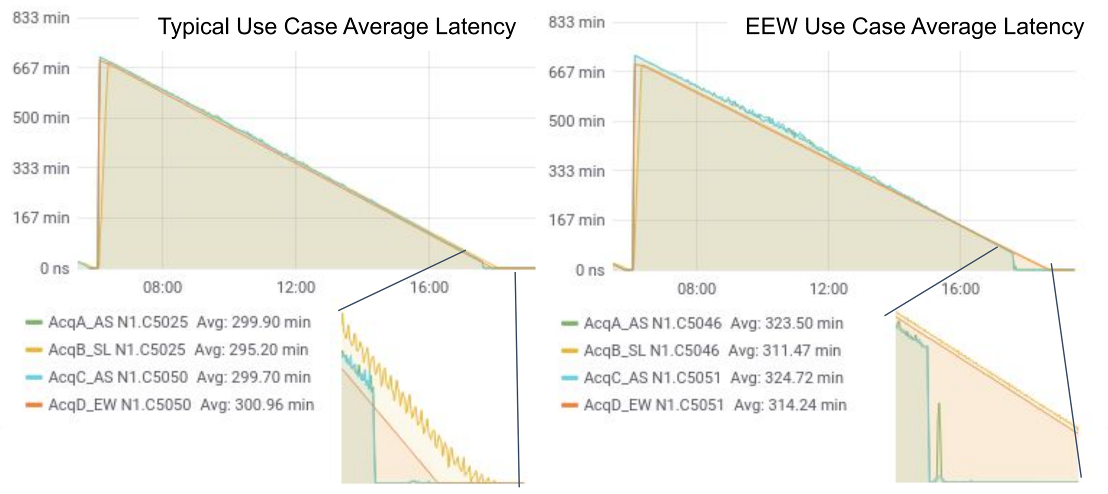
- Bandwidth utilization is generally consistent with baselines, with some variation as expected.
- Latency generally follows the expected pattern for the outage recovery. On reconnect, the SL and EW systems hold back real-time data pending recovery of the outage, resulting in high latencies for over an hour in comparison to the NP-acquired data. On bandwidth constrained links, this delay could be significantly longer.
- All systems have 100% data completeness with the exception of EW. The EW archive had gaps and / or overlaps on all channels.

Bandwidth

Typical Use Case Total Bandwidth:
- AcqA: 3560 bps, AcqB: 3874 bps
- AcqC: 3566 bps, AcqD: 3691 bps

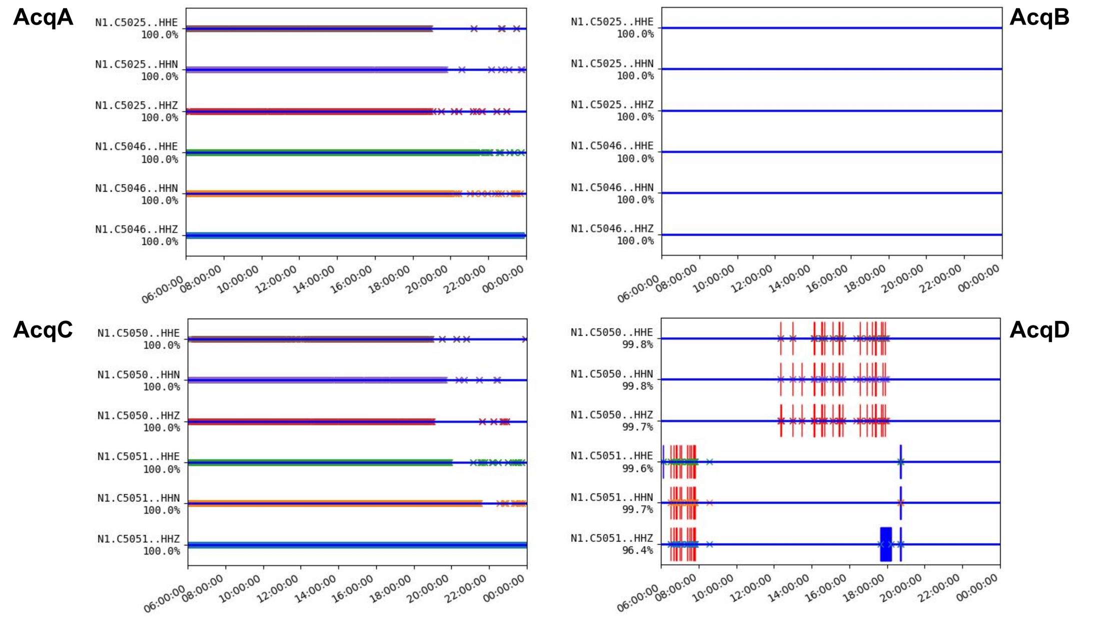
EEW Use Case Total Bandwidth:
- AcqA: 8009 bps, AcqB: 8701.5 bps
- AcqC: 7984 bps, AcqD: 17614 bps

Latency



The insets show that once re-connected, latency of NP stations immediately returns to normal, while Seedlink prioritizes recovery of missed data first.

Data Completeness



FEATURE REVIEW

NP:

- **Real-time prioritization:** NP prioritizes real-time data ahead of filling gaps, so it can immediately start streaming live data, while recovering gaps when bandwidth allows after an outage. This gives the lowest possible overall latency and fulfils requirements for early-warning systems.
- **Gap recovery:** NP will continue to work to retrieve missed data until it is recovered or known to no longer be available at the source.
- **Multicast streaming:** NP supports multicast streaming, which allows multiple destinations to receive the same data stream, supporting data center redundancy without additional bandwidth.
- **Short-term complete:** NP feature to ensure clients which require data to be in order only receive ordered data while allowing ongoing gap recovery to ensure maximum data completeness.
- **Throttling:** NP feature to limit bandwidth utilization to a preset maximum.
- **Fragmentation:** NP feature to limit the maximum packet size, avoiding the need for IP Fragmentation, which is often not supported in routers.
- **GNSS support:** NP supports transport of arbitrary payloads, including GNSS BINEX data, allowing unified acquisition systems.
- **SOH support:** NP has a robust, self-describing and extensible SOH system to allow for intelligent network monitoring, without assuming units and other pertinent information of the SOH data.

Seedlink:

- **In-Order:** By using TCP, data is guaranteed to be in-order, removing need for short-term complete, but at the cost of higher data usage and lack of live data prioritization.
- Seedlink/TCP dynamically adjusts to use maximum available bandwidth.
- Simple state in receiver, storing just the last sequence number received.

CONCLUSIONS

For typical digitizer use cases, with reliable telemetry, there is little that differentiates SeedLink and NP. For early warning applications, with reduced packet durations to minimize latency, NP is much more bandwidth efficient, requiring less than half the total bandwidth of SeedLink.

NP generally produced lower latencies than SeedLink across the various use cases and conditions. A contributing factor is that NP prioritizes real-time data over gap recovery.

In general, data completeness for NP-acquired data was found to align with or exceed that of SeedLink systems, especially when considering that NP systems will continue working to recover old gaps after connectivity has been restored.

NP is a full-featured protocol, which can be optimized to provide efficient and robust acquisition for many different use cases. Multicast streaming allows data center redundancy without additional bandwidth. Short-term complete can ensure clients receive data in order. Throttling provides direct control over bandwidth. Encapsulation of other formats, like BINEX, allows GNSS and other systems to fully leverage these benefits.

For unreliable and / or bandwidth-constrained telemetry links where real-time acquisition latency is a priority, the NP protocol provides an efficient and robust data acquisition solution.

NEXT STEPS

- Explore potential sources of latency discrepancies and gaps
- Characterize directional bandwidth utilization
- Experiment with different acquisition system parameters
- Expand study to include other digitizer models and acquisition protocols
- Expand study to include GNSS receivers and GNSS data acquisition systems