HIGH PERFORMANCE COMPUTING FOR GEOSPATIAL, VEHICULAR, AND NETWORK DATA ANALYSIS

Project Description

Growing dependence on digital communications and the availability of highly sophisticated communications systems require more complex and detailed test events, producing terabytes of data per day. When aggregated over time, the processing requirements are beyond the scope of commodity computational resources or even small clusters. Scalable, parallel methods are required to transfer and process data into a form that analysts can use in a timely manner.

Prior to using High Performance Computing (HPC) assets to process test data, a Java-based application was executed on several large (64 core) Linux servers to process collected data. This could involve several days of processing for a single day's collection set. The design goal for a new, HPC-based process was to complete all data processing and provide the results within several hours. To accomplish this order-of-magnitude reduction in time, the Aberdeen Test Center (ATC) partnered with ARL to develop a prototype HPC framework and created a new automated data transfer and job control pipeline.

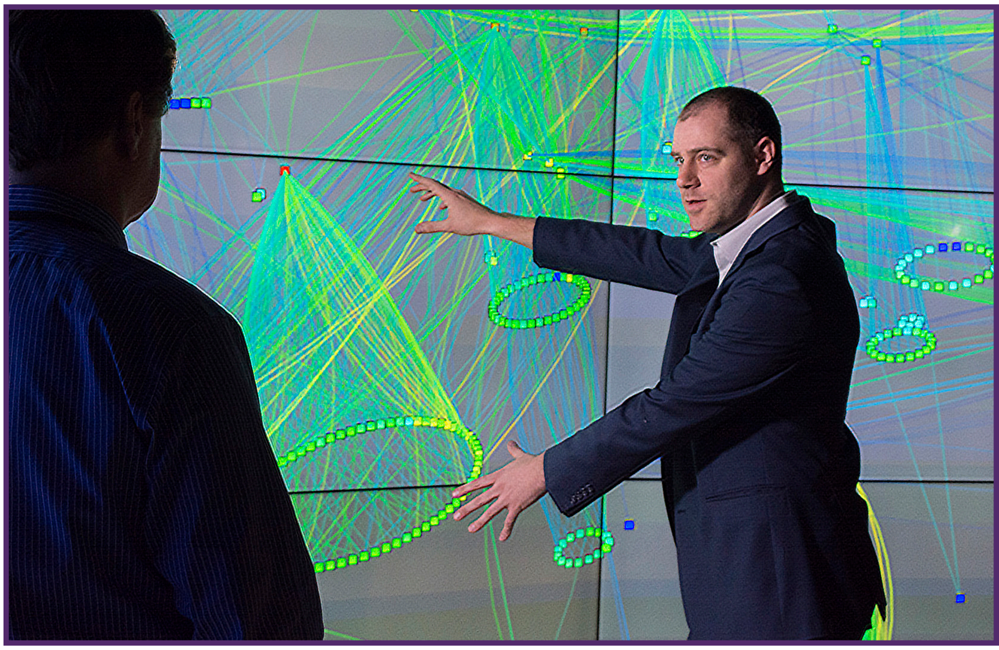
Relevance of Work to DOD

Tactical networks support the activities of a military unit during operations, with high assurance and minimal delays, as the unit maneuvers to accomplish its mission. Vehicles are also becoming more complex with Controller Area Networks (CAN) buses that support vehicle operation and status monitoring. Large-scale tests are required to provide senior leaders with data necessary for system evaluation and milestone decisions. The software developed provides scalable, parallel methods, which transfer data from test ranges and process it into a form that analysts and decision-makers can use in a timely manner. An example of this data is shown in Figure 1, which shows temporally controlled geospatially rendered satellite (blue) and terrestrial (green) communications links between tactical vehicles.

Computational Approach

An automated event-driven pipeline (depicted in Figure 2) was created to transfer data, marshal HPC jobs, and load results around the clock for multiple data sets simultaneously; it enabled decoupling distributed HPC processing from the results database.

A framework was designed to distribute tasks in multiple phases over an arbitrary number of processing cores. Input files are read once per phase, and multiple modules "subscribe" to data record types and receive a copy of the data record for processing. The framework consists of many data-processing modules, grouped and executed in separate sequential phases based on data dependencies. Early phases are typically dependent on only raw data, while later phases typically rely on only the output of the early phases. Data exchanged between phases is typically performed via files. While this approach increases input-output (I/O) requirements, it is essential for the orthogonal redistribution of data between phases instead of exchanging the data being processed. This improved reliability of the data path and allowed for check pointing with little effect on execution speed.
The framework consists of many data-processing modules, grouped and executed in separate sequential phases based on data dependencies. Each processing



module describes module dependencies and the framework automatically schedules a processing module in the earliest phase possible. Early phases are typically dependent on only the original data collected in the field, while later phases typically rely on only the output of an earlier phase. Data exchanged between phases is typically performed via HDF5 data sets. Though this approach increases I/O requirements, it is essential for the orthogonal redistribution of data between phases.

Results

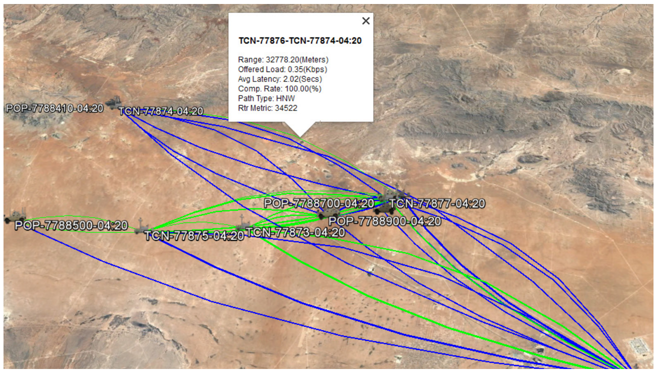
The HPC framework produces results with an order of magnitude reduction in time compared to previous methods. The results are automatically loaded in a relational database schema that enables analysts and other customers to combine data from multiple sources for determining system reliability and performance.

The initial database schema included geospatial, network, and device status data. It was expanded to include CAN bus data dynamically identified during HPC processing. This flexibility supports raw CAN bus messages, J1939 and MIL-STD-1553 specifications and network traffic containing ASN.1 payloads.

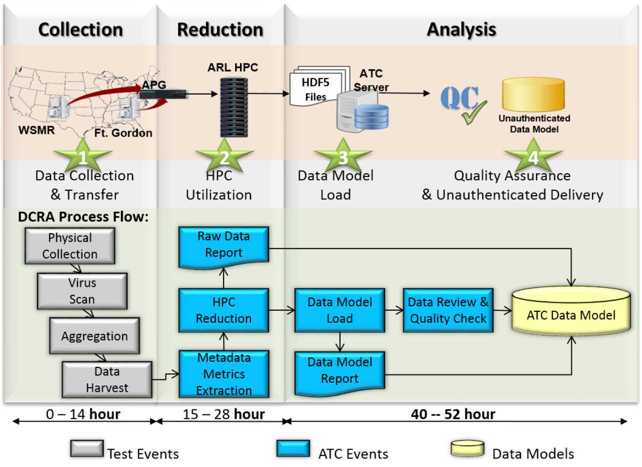
Figure 3 displays new features added to the HPC processing. This figure shows the state of each vehicle's Coordinated Adaptive Cruise Control (CACC) state, which is derived from CAN bus observations in the vehicle. The HPC processing combined CAN, GPS and Dedicated Short Range Communications data to produce the data product showing the behavior of the CACC software under dynamic conditions.

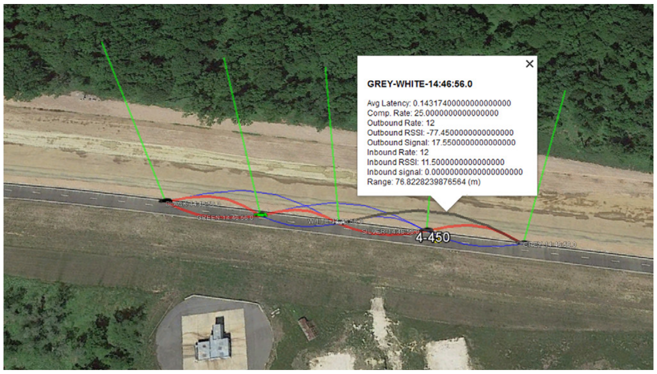
Future

Future research will replace the current HPC code and database with an open source Apache Hadoop ecosystem. Hadoop provides a distributed file system and query engine that will alleviate I/O and data movement issues; the processing and querying both run on the same distributed HPC platform. Queries will also gain the speedup of this distributed system. Web front-ends and interfaces will enhance analysis and expedite decision-making.


TCN-77876-TCN-77874-04:20

Range: 32778.20(Meters)
Offered Load: 0.35(Kbps)
Avg Latency: 2.02(Secs)
Comp. Rate: 100.00(%)
Path Type: HNW
Rtr Metric: 34522




GREY-WHITE-14:46:56.0

Avg Latency: 0.14317400000000000000
Comp. Rate: 25.0000000000000000
Outbound Rate: 12
Outbound RSSI: -77.4500000000000000
Outbound Signal: 17.5500000000000000
Inbound Rate: 12
Inbound RSSI: 11.5000000000000000
Inbound signal: 0.00000000000000000000
Range: 76.8228239876564 (m)

Figure 1: (top) Geospatial visualization of communication links in mobile tactical network.
Figure 2: (middle) Automated data collection, reduction, and analysis pipeline.
Figure 3: (bottom) Department of Transportation Coordinated Adaptive Cruise Control Testing at ATC.

Contact Information

James Feight
US Army Aberdeen Test Center
Test Technology Directorate
james.r.feight.civ@mail.mil 410.278.9155

Co-Investigators

Brendan Tauras, James Adametz, & Brian Ramsel
US Army Aberdeen Test Center

Brian Panneton
Army Research Laboratory
Computational and Information Sciences Directorate