ARL-TR-7409 ● SEP 2015

US Army Research Laboratory

High-Bandwidth Tactical-Network Data Analysis in a High-Performance-Computing (HPC) Environment: Introduction

by Kenneth D Renard, James R Feight, Michael Amabile, and James Adametz

NOTICES

Disclaimers

ARL-TR-7409 ● SEP 2015

US Army Research Laboratory

High-Bandwidth Tactical-Network Data Analysis in a High-Performance-Computing (HPC) Environment: Introduction

by Kenneth D Renard
Computational and Information Sciences Directorate, ARL

James R Feight
Aberdeen Test Center, Aberdeen Proving Ground, MD

Michael Amabile
Focused Ingenuity LLC, Bel Air, MD

James Adametz
QED Systems, LLC, Aberdeen, MD

REPORT DOCUMENTATION PAGE					*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*
September 2015

2. REPORT TYPE
Final

3. DATES COVERED (From - To)
1 July 2012–31 December 2014

4. TITLE AND SUBTITLE
High-Bandwidth Tactical-Network Data Analysis in a High-Performance-Computing (HPC) Environment: Introduction

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Kenneth D Renard, James R Feight, Michael Amabile, and James Adametz

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Army Research Laboratory
ATTN: RDRL-CIH-C
Aberdeen Proving Ground, MD 21005-5067

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-7409

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Over the past 2 years the US Army Test and Evaluation Command, Aberdeen Test Center, has worked in close collaboration with the US Army Research Laboratory's Computational and Information Sciences Directorate to utilize high-performance computing to build a system-of-processes to reduce large volumes of data recorded in a high-bandwidth tactical network test environment. This introductory report will describe the nature of a tactical ad-hoc network, how such an aggregate is tested in field events, and the types of refined products and analysis performed on collected data. The report will also introduce a series of reports written about this system-of-processes, in which each report describes in more detail specific aspects of the processing.

15. SUBJECT TERMS
tactical networks, data reduction, high-performance computing, data analysis, big data

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Kenneth Renard
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include area code)
Unclassified	Unclassified	Unclassified	UU	20	410-278-4678

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

1. Introduction

Tactical networks support the activities of a military unit during operations. This support includes the ability to transport voice and data messages, with high assurance and minimal delays, as the unit maneuvers to accomplish its mission. Tactical networks for the US Army have evolved significantly in recent years, transitioning from low-capacity, fixed-configuration radio systems to Internet Protocol networks comprising tiered high-capacity transmission systems linked by ad-hoc routing protocols (see Fig. 1). This evolution significantly increases the scale and complexity of testing such networks, from the component level through the integrated "tactical internet". Within the Army, the standard unit of testing in recent events has been the brigade, which includes approximately 60 "upper tier" nodes with high-capacity connections and nearly 1,000 "lower-tier" nodes with less-capable connections linked via the upper tier. The additional network complexity within each tier has also significantly increased the data collection requirements levied by program offices and evaluation agencies, who require detailed, multivariate products to conduct their analysis efforts. The combination of increased scale and complexity leads to raw data sets on the order of 1 TB per day. When aggregated over a multiday or multiweek test effort, the processing requirements for such raw data sets are beyond the scope of commodity computational resources or even small clusters.

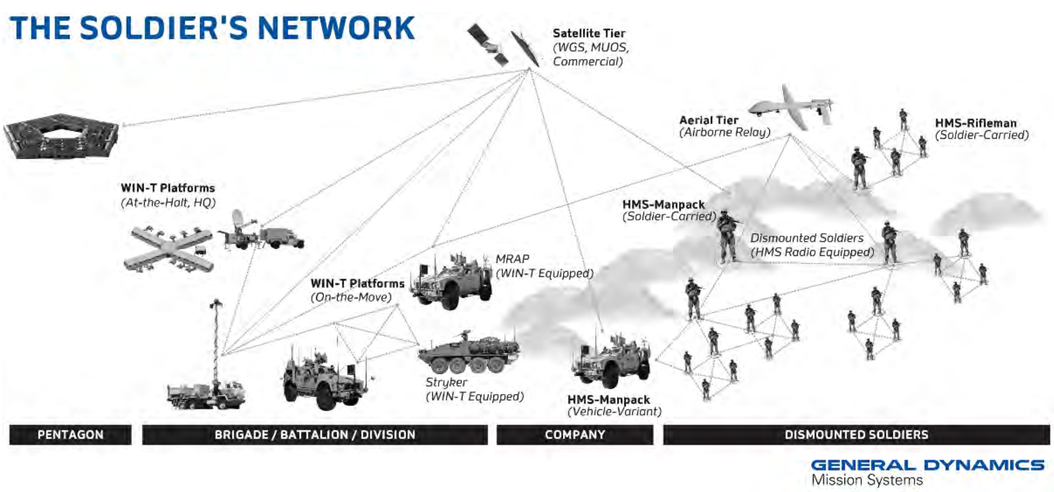

Fig. 1 Tactical networking[1](image used with permission from General Dynamics)

2. Complexities of Administrating High-Profiled Event

Today's rapidly evolving global environment requires an agile Army capable of conducting multiple large-scale efforts simultaneously around the world. As a result, the Army is continuously developing new material solutions and upgrades for existing systems to ensure battlefield superiority. The US Army Aberdeen Test Center (ATC), a Major Range and Test Facility Base subordinate to the US Army Test and Evaluation Command (ATEC), is dedicated to providing unbiased test and evaluation (T&E) products to assess new materiel before it is fielded to Soldiers.

When planning for or executing any test effort, ATC works closely with evaluators, testers, and vendors to determine the best and least intrusive way to collect and reduce data to a consumable format needed for analysis. Test capabilities and processes must evolve with the community to ensure Warfighter dominance in the field. ATC has become agile in addressing emerging fiscal and strategic guidance while ensuring a qualified workforce to achieve successful test efforts.

Large-scale tests provide the T&E community with a number of valuable opportunities, including increasing amounts data for evaluations, revitalizing development of a system, and providing an increased sense of community. With these benefits comes greater responsibility for local test centers to ensure they work closely with program managers and evaluators to collect, reduce, and analyze data correctly. During large-scale tests, such as Network Integration Exercise 14.2, the potential existed for days of testing being lost due to unusable data, discussions between testers and evaluators, and added risk to program success. Therefore, it was important for testers to actively engage in strategic and collaborative planning, which involved a community effort. Test centers should develop an overall philosophy and define the specific tests and strategies they will use to collect, reduce, and analyze data. The evaluators should verify that these plans coincide with their needs for their evaluation. The program office should also verify that the tests performed will fully test their system and provide unbiased data. Above all of these considerations, the test centers need to ensure the safety of Soldiers, data collectors, testers, and support personnel on site executing the test.

To achieve these cooperative goals, ATC has focused effort in 5 areas of growth:

- **Partnership:** Forge mutually beneficial alliances across the Army and the Department of Defense (DOD) to understand how to test in the future.
- **Product:** Ensure the proper capabilities are available and cost-effective and the resulting test products are timely, succinct, and defendable.
- **Process:** Identify areas to gain efficiencies in testing and ensure all testing processes are standardized and trackable.

- **Purse:** Keep abreast of all emerging fiscal guidance to ensure test capabilities remain efficient, sustainable, and mission-focused.
- **People:** Continually grow and develop of a qualified, flexible, accountable mission-centered workforce.

Focusing on these areas will help test officers maximize their efforts by learning from other agencies and adopting proven practices. Too often, however, past lessons learned are not documented in a clear and concise manner. To address this information gap, ATC has partnered with the US Army Research Laboratory (ARL) to develop a series of technical reports to document the efforts and process created and executed during large-scale tests.

ATC has maintained a legacy of excellence in testing for almost a century by committing to these areas of growth. ATC's diverse test mission has allowed them to remain viable as the threats to the nation have evolved. ATC continually reassess its organization, procedures, and culture of supporting customers on every test to ensure they can address the tests and customer's requirements and needs. Diligent partnership with other test centers, evaluators, and program offices have helped identify areas of growth, ensuring T&E capabilities are matured and ready to test emerging Warfighter technologies. As ATC goes further into the future, the testing community must develop creative funding solutions, embrace change, and help lead the Army in data collection and testing efficiency. The community must heavily invest in its workforce and empower them to be technically proficient leaders who understand their role in the Army profession. Sticking to this plan has provided a cohesive roadmap allowing ATC to be a premier test center within the DOD. It keeps in step with Army values while allowing ATC to be responsive, flexible, and strategically positioned to provide superior service to ATEC, the Army, and the DOD.

3. Current State of High-Performance Computing

Significant advancements have been made in the past decade in building faster and more capable computing hardware. General purpose computers today are orders of magnitude faster than comparable machines from 10 years ago. A similar growth has occurred in high-performance-computing (HPC) systems, allowing larger and more complex computing problems to be solved. Today's HPC platforms typically combine many high-end traditional computing components into a "cluster" of systems that can work together to solve large problems. Where a typical desktop today may have a multicore central processing unit (CPU), an HPC cluster may have hundreds of thousands of CPU cores connected by high-speed, low-latency links that allow these cores to work closely together on the same task. The aggregate memory capacity of these systems enables the computation of problems much larger than could be executed on a high-end workstation. This hardware must be combined with specialized software capable of utilizing the set of resources in an efficient manner.

Algorithms that process and reduce raw data must be “parallelized” to correctly and efficiently break, or partition, a larger problem into smaller sets of individual problems such that they can then be executed independently and concurrently by separate processors. The results of these smaller problems are typically combined to provide an aggregate result or are further partitioned to process another type of result. The concept of “map-reduce” (also known as “scatter-gather”) has been employed historically in the world of distributed computing. The goal is to map, or assign, parts of a large data set to many execution tasks for initial processing of the data. The reduction phase gathers the results from each execution task and combines them into a single set of results.

Complex problems typically require several rounds of mapping and reduction, where data is partitioned in different, orthogonal ways for each round. Careful decomposition of the data or problem needs to be developed so that the integrity of the solution is not compromised. Communication of data between rounds of map-reduce are optimized to require minimal bandwidth and time. Algorithms for decomposition and processing are coded in software and executed on HPC machines.

The HPC resources used in this effort are provided by the DOD High Performance Computing Modernization Program (HPCMP). A DOD Supercomputing Resource Center (DSRC) at ARL houses and maintains such resources for use by the DOD community. Users of the HPC systems submit jobs to shared resources, which are queued for execution. As computing resources become available, a set of CPUs and memory from the cluster becomes allocated for the job to use. Depending on the amount of resources required for the job, wait times can vary from seconds to days. The user specifies how many CPU cores are required for their job and an expected maximum run time. The HPC system's scheduler resolves these requirements along with those of other users and manages the starting and stopping of jobs. Jobs that take more time or memory than were originally requested are terminated.

Significant time and expertise are required to engineer software that integrates into the HPC environment to realize the power of distributed computing. For significantly large and complex problems, this is essential and pays huge dividends in the ability to obtain results in greatly reduced timeframes.

4. Data Reduction Timeline Issue

Prior to using HPC assets to reduce field test data, an ATC-developed Java-based application was executed on several large (64 core) Linux servers to process collected data. This process could involve several days of processing for a single day's collection set, even when multiple copies of this application were executed in parallel.

The design goal for HPC-based processing was to complete all data processing and provide the results within several hours after initiating a run.

To accomplish this nearly order-of-magnitude reduction in processing time, ATC partnered with ARL to leverage HPC computing resources and adapt existing algorithms for use in a massively parallel environment. ATC staff provided subject matter expertise in the form and structure of the input data, and the current algorithms employed for reduction. ARL staff then developed a prototype framework that suited the data reduction problem and shared their expertise in HPC software design. Collaboratively, the framework became populated with specific data-processing modules that generated the target data products for the analysis community. Primary authorship of processing modules came from both organizations as did additional code contributions.

5. Tactical Network Testing and Data Collection

The DOD acquisition process calls for 2 main types of testing: developmental testing, in which the environment is controlled and the purpose is to vet the ability of the system to meet its technical requirements, and operational testing, in which the system is placed into the hands of Soldier role players, who attempt to employ the system within the context of an assigned mission

For Army tactical networks, the technical data collection requirements for both types of testing were developed in parallel. A "multimodal" approach was designed and implemented, which involves the simultaneous collection of multiple data products from various points within the integrated network to inform a more complete analysis. Given the complexity of these networks, it is generally not possible to assess performance or requirement satisfaction using single data products or metrics.

This multimodal approach includes the following key elements:

- Packet Capture: The collection of network traffic offered by user applications at source and destination nodes, which is used to calculate the load applied to and delivered by the network and the losses and delays imposed by the network.

- Time Space Position Information: The collection of the location of network nodes over time, which is used to calculate relative geometry and mobility state; both are key factors in understanding the performance of networks in motion, where links break and form in an ad-hoc manner.
- Subsystem Statistics: The collection of state information from key system components, including radios, modems, routers, switches, and traffic handling systems. Such information is generally captured via Simple Network Management Protocol (SNMP) polling or database queries and informs the state of such devices over time and the state of the radio or routing links between them.
- System Configuration: The collection of the state of the configuration of the system over time. Within modern tactical networks, performance can be driven as much by the capability of the components as the configuration that is applied to them. Users have the ability to apply configurations, which can help or hinder overall performance, and it is often not possible to properly interpret data products without knowledge of the running configuration. This configuration is also key in the postprocessing of the other data products, since it enables proper association of the independently collected products.

This collection approach is applied locally across all nodes of the network; independent instruments are installed on key nodes, configured to collect per a self-consistent design, and operate on a noninterference basis during a test event. After each event, or periodically during a long-running event, data is "harvested" from the distributed collectors and aggregated at a single location for processing. Here, a series of complex algorithms organizes, inventories, and applies correlation and other reduction processes to the raw data products to form an integrated product usable by analysts. Typically, the integrated product is realized as a database schema that contains elements of the raw data along with calculated metrics specified by the program office or evaluators.

6. Types of Analysis

Analysis efforts within the DOD acquisition process are of 2 basic types: requirements validation, where the goal is to assess a specific metric (or set of metrics) under carefully controlled conditions, generally as specified in a requirements document or system specification, and system characterization, where the goal is to control one or more "factors" (independent variables) and record the "response" (dependent variables).

In both cases, the multimodal collection approach is generally required to properly assess modern tactical networks. Key elements of analysis generally include the following:

Traffic Performance Statistics. These provide data on the offered load, the rate of "completion" (i.e., delivery) of that load, and the time required for delivery. These are typically key metrics for networks and drive most performance requirements. Figure 2 is a typical way in which load and completion rate are portrayed temporally for a typical network link.

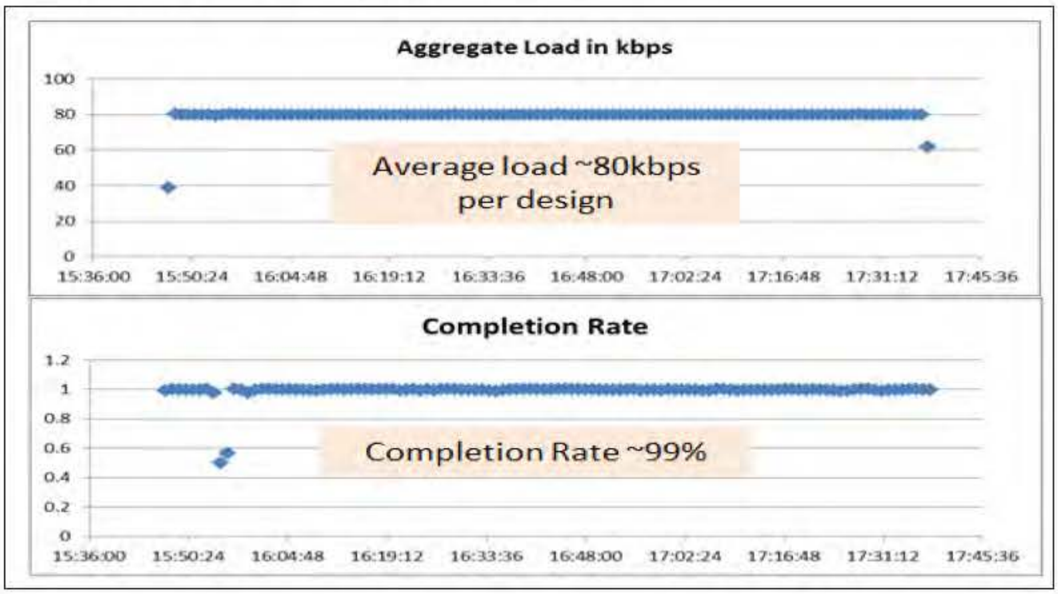

Fig. 2 Key network metrics

Higher-Level Traffic Metrics. These are also key in assessing the ability of the network to not just deliver traffic but to support the semantics of complex application flows. Statistics on Transmission Control Protocol (TCP) are critical to such analysis efforts, in addition to metrics drawn from application-level interactions, such as Voice over Internet Protocol (VoIP) signaling and messaging formats such as Joint Variable Message Format (JVMF). Figure 3 illustrates statistics relevant to the usage of TCP by a currently fielded Army application, including session distribution over time and the session start-up delay.

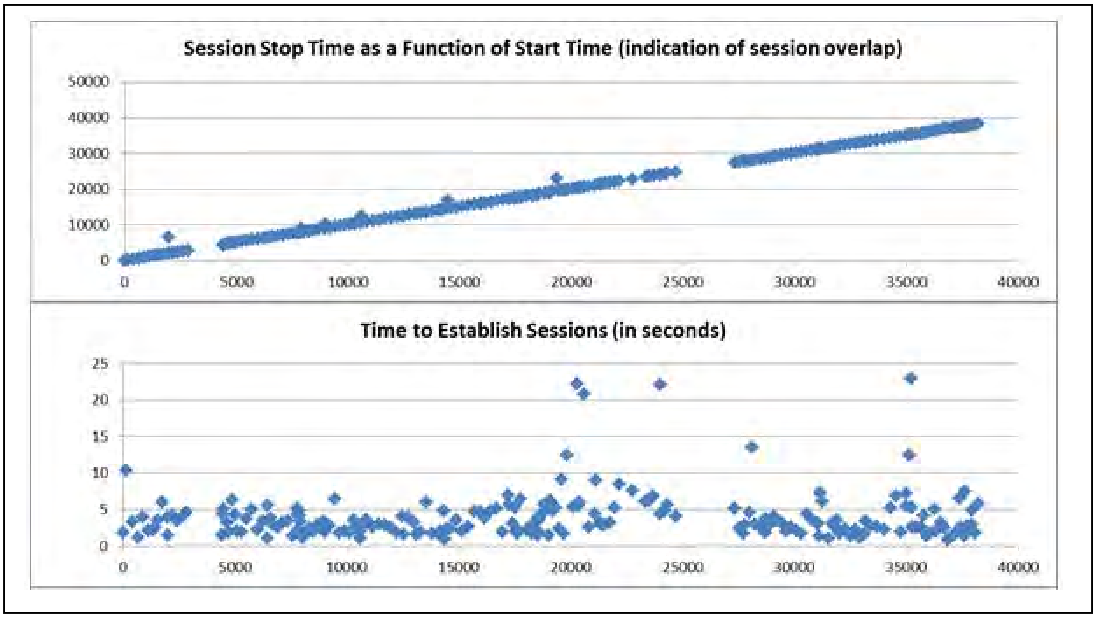

Fig. 3 Higher-level traffic metrics (TCP)

Network Performance Statistics. These provide information on the state of IP routing tables and radio-level connections, which informs the overall state of the network and complements the view provided by the traffic performance statistics. Often, within the characterization of the network, the juxtaposition of connectivity and traffic data is critical in understanding performance limitations. Figure 4 illustrates the completion rate of traffic between 2 maneuvering nodes as the route through which the traffic flows switches between terrestrial line-of-sight links and satellite communication links; such a view would not be possible using a single collection modality.

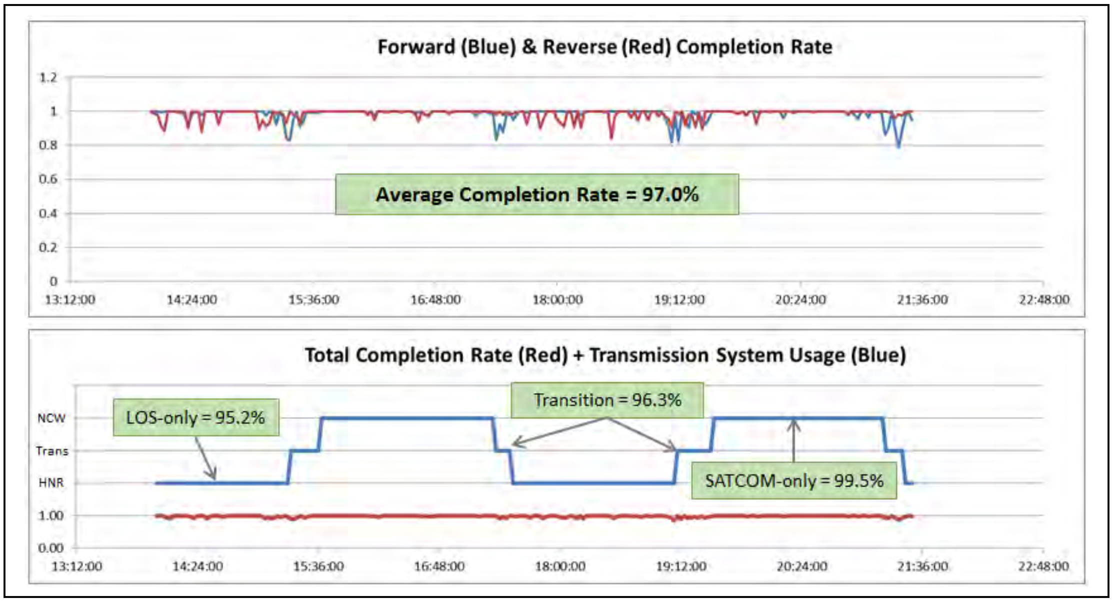

Fig. 4 Completion rates between nodes

ATC's customer organizations include Program Offices, the Army Evaluation Center (AEC), and oversight offices from the Office of the Secretary of Defense. These organizations typically employ the HPC-produced database schema to generate such analysis products that provide quantitative material for their reporting activities. This schema was developed in concert with the customer organizations and has grown over time to include dozens of relational tables containing hundreds of data elements.

7. Further Reading

As mentioned earlier, several reports related to this introductory report have been published that provide more detailed views into how ATC and ARL worked together to build the system of processes that enabled the team to successfully process on average 1.5 TB of data within the intended 72-h reduction timeframe.

The first of these reports[2] addresses the issue of marshalling the data from the field all the way to the endpoint of populating an authenticated data model used by the evaluators and analysts.

The next report[3] explores the HPC framework that was employed to reduce the raw data. Closely related to this framework report is a detailed explanation of the complex processing required to handle distributed time tagging of the data,[4] which is crucial for accurate calculation of network loads, completion rates, and latency data.

Two of the reports[5,6] cover how the low-level network communications processing (at the IP and TCP levels) was accomplished.

Processing of the data at the application level is presented in Renard et al.[7] This report is focused on VoIP and JVMF command and control application message processing.

The last of these related reports explores the processing and handling of the SNMP, and Network Management System polled data are given.[8]

One additional relevant reading would be ATC's *Technical Report: Collection, Reduction & Analysis of Data from the Warfighter Information Network – Tactical Increment 2 Network*. This report is a good example of how the data rendered by the system-of-systems presented in this report can be used to analyze the performance of a complex system under test.[9]

8. Conclusion

The collaborative efforts between the Analysis team at ATC and the HPC team at ARL over approximately 2 years have resulted in a system of processes that successfully reduced more than a terabyte of high-bandwidth network traffic per day.

Additional collaborative efforts between ATC, AEC, Project Manager WIN-T, the Director of Operational Test and Evaluation, and several other analysis organizations ensured that the final product produced by the reduction system provided the quantity and quality of data needed to properly evaluate the systems under test in a time period that enabled the analytical communities ample time to render a final report on system performance.

9. References

1. General Dynamics. The Soldiers Network. 2014 [accessed 2015 Jan 15]. http://thesoldiersnetwork.com/wp-content/uploads/2014/10/The-Soldiers-Network-Diagram1-e1414769652154-1024x424.png.

2. Renard K, Rivera JD, Adametz J, Franssen J. High-bandwidth tactical-network data analysis in a high-performance-computing (HPC) environment: data marshalling. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2015 Sep. Report No.: ARL-TR-7410.

3. Panneton B, Adametz J. High-bandwidth tactical-network data analysis in a high-performance-computing (HPC) environment: hpc data reduction framework. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2015 Sep. Report No.: ARL-CR-0777.

4. Panneton B, Adametz J, Franssen J. High-bandwidth tactical-network data analysis in a high-performance-computing (HPC) environment: time tagging the data. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2015 Sep. Report No.: ARL-CR-0778.

5. Panneton B, Adametz J . High-bandwidth tactical-network data analysis in a high-performance-computing (HPC) environment: packet-level analysis. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2015 Sep. Report No.: ARL-CR-0779.

6. Renard K, Adametz J. High-bandwidth tactical-network data analysis in a high-performance-computing (HPC) environment: transport protocol (transmission control protocol/user datagram protocol [TCP/UDP]) analysis. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2015 Sep. Report No.: ARL-TR-7411.

7. Renard K, Besack G, Dixon D. High-bandwidth tactical-network data analysis in a high-performance-computing (HPC) environment: voice call analysis. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2015 Sep. Report No.: ARL-TR-7412.

8. Panneton B, Tauras B, Wancowicz C. High-bandwidth tactical-network data analysis in a high-performance-computing (HPC) environment: device status data. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2015 Sep. Report No.: ARL-CR-0780.

9. Aberdeen Test Center. Technical report: collection, reduction & analysis of data from the warfighter information network – tactical increment 2 network. Aberdeen Proving Ground (MD): Army Research Laboratory (US); Aberdeen Test Center; 2015 Apr. (FOUO).

List of Symbols, Abbreviations, and Acronyms

AEC	US Army Evaluation Center
ARL	US Army Research Laboratory
ATC	US Army Aberdeen Test Center
ATEC	US Army Test and Evaluation Command
CPU	central processing unit
DOD	Department of Defense
HPC	high-performance computing
IP	Internet Protocol
JVMF	Joint Variable Message Format
SNMP	Simple Network Management Protocol
T&E	test and evaluation
TCP	Transmission Control Protocol
VoIP	Voice over Internet Protocol
WIN-T	Warfighter Information Network - Tactical

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
2 (PDF)	DIRECTOR US ARMY RESEARCH LAB RDRL CIO LL IMAL HRA MAIL & RECORDS MGMT
1 (PDF)	GOVT PRINTG OFC A MALHOTRA
1 (PDF)	DIR USARL RDRL CIH C K RENARD