AI/ML-driven Network Optimization to enable Synthetic Training and Distributed Simulation

Gregg Patti, Ebony Robinson, Jack Burbank, June Gordon, Todd Lutton, Antonio Fiuza
Sabre Systems, Inc.
Warminster, PA
{gpatti, erobinson, jburbank, jgordon, tlutton, afiuza}@sabresystems.com

Brad Friedman
U.S. Army Futures Command
Orlando, FL
brad.d.friedman.civ@army.mil

ABSTRACT

Any federated simulation environment or synthetic training environment is only as effective as the network that interconnects their various components. A poor performing network limits the ability for real-time or near real-time distributed simulation and reduces the effectiveness of the synthetic training experience. Low-capacity networks constrain data volume that can be shared throughout the environment. Low performing networks lead to unacceptable latency that translates to a poor user experience. The network must reliably provide the capacity and performance that enables these key simulation and training paradigms. Network performance becomes more vital as data intensive technologies such as Augmented Reality / Virtual Reality (AR/VR) are introduced to the training environment.

This paper presents the High-Performance Synthetic Training Environment (STE) Optimization Network (HP-SON), a capability currently being developed by Sabre Systems, Inc. for the US Army Futures Command Synthetic Training Environment (STE) Cross Functional Team (CFT). To address the goal to inter-connect a wide range of geographically distributed simulation and training elements, HP-SON incorporates a suite of artificial intelligence / machine learning (AI/ML) algorithms to optimize the various transport networks and enable the Army's full large-scale, distributed vision of the STE. HP-SON's AI/ML employs state-of-the-art AI/ML algorithms dynamically learn network performance and user activity, resulting in a highly-reliable set of predictive network analytics. These predictive network analytics are then used to drive several predictive AI/ML-driven features, including data compression, traffic scheduling, protocol optimization, and forward data caching. These capabilities work together to improve perceived network performance and capacity to produce an improved user experience.

ABOUT THE AUTHORS

Gregg Patti is a Software Development Engineer at Sabre Systems. Mr. Patti earned his Bachelor of Science in Computer Information Systems, with additional Minors in Computer Science, Networking & Cyber Security, and Game Design & Development, from SUNY Polytechnic Institute in 2022. Mr. Patti is familiar with several fields of software engineering, thanks to his wide range of studies.

Ebony Robinson is a Software Developer at Sabre Systems. Ms. Robinson is proficient in several programming languages, development tools, and techniques. She possesses a basic understanding of cyber security principles and concepts and is eagerly improving her knowledge in the field. Ms. Robinson earned her Bachelor of Science degree in Computer Science with a specialization in Software Engineering from Radford University in 2021.

Jack Burbank is the Executive Director of Advanced Communications Technology at Sabre Systems where he designs, develops, and evaluates next-generation wireless systems and network technologies. Mr. Burbank is an expert in the areas of wireless networking, modeling and simulation, wireless system development, and wireless network security. Mr. Burbank has published over 50 technical papers on topics of wireless networking and has contributed to multiple books related to wireless networking. Mr. Burbank has authored books on the topics of Wireless Networking and Modeling and Simulation. Mr. Burbank is co-editor of the Wiley-IEEE Press book series on IEEE standards. Mr. Burbank earned his Bachelor of Science and Master of Science degrees in Electrical Engineering from North Carolina State University in 1994 and 1998, respectively, and is currently a PhD student at the University of North Dakota.

June Gordon has many years of experience providing systems engineering, software development, and mathematical modeling services to both government and industry. She has been the Principal Investigator on numerous research projects, including multiple Small Business Innovative Research (SBIR) initiatives. Her areas of expertise include development of advanced acoustic and non-acoustic signal processing capabilities, target tracking and data fusion methodologies, resource allocation and optimization algorithms, large-scale simulation models, environmental and atmospheric models, real-time command and control software, and deep learning models. Ms. Gordon has a B.A. in Mathematics from Temple University and an M.S. in Applied Mathematics from Drexel University.

Todd Lutton is a Program Manager supporting the Cryptographic Modernization Branch, CSIA Division, Engineering & Systems Integration Directorate, DEVCOM C5ISR Center at Aberdeen Proving Ground, where he is part of a team that evaluates technologies and their possible application to improve Army network or device security. Mr. Lutton is a co-inventor on two Army patents related to a Multi-Function Smart Communications Card. Mr. Lutton earned his Bachelor of Science degree in Systems Engineering and Applied Science / Computer Science from the United States Naval Academy in 1984.

Antonio Fiuza Antonio Fiuza is the Director, C4I Science & Technology Programs at Sabre Systems and currently is the Program Manager for the HP-SON SBIR. Mr. Fiuza started with Sabre in 2011 and provided consulting services to the Space & Terrestrial Communications Directorate, CERDEC, Aberdeen Proving Ground, MD for the conceiving, marketing and defending the complex R&D communications, networking and cyber security programs. Includes the core R&D program, strategic planning, Small Business Innovative Research (SBIR), Small Business Technology Transfer (STTR), special studies and international programs. He also acted as the Executive Chair of the Joint Tactical Edge Networking Group, a joint committee designed to coordinate service communications and networking R&D across the OSD community. Previously, Mr. Fiuza was the Associate Director for Technology, Space & Terrestrial Communications Directorate Company at the U.S. Army Communications Electronics Research, Development & Engineering Center (CERDEC), Ft. Monmouth, NJ. He led the long range strategic technical planning in the Directorate, revamping how the organization conducted strategic planning in all phases. This included managing the >$80M Science & Technology Base and R&D program for the Directorate and leading efforts to formulate new Science & Technology Programs, allocate funds, and coordinate with other Directorates and outside agencies. Execute the Directorate outreach program to include Joint Projects with the Air Force/Navy, DARPA, international and formulation of partnerships with Industry/Academia. Chaired the Tri Service/Defense Advanced Research Projects Agency (DARPA) Technology Focus Team for Reliance 21. Mr. Fiuza is a graduate or the New Jersey institute of Technology, BSEE 1980.

Brad Friedman currently leads the computing efforts for the Army Synthetic Training Environment (STE) Cross Functional Team (CFT) and has 20 years of experience as an engineer, scientist, test evaluator, and program manager supporting DISA, MARCORSYSCOM, Army NETCOM and AFC. He has occupied significant roles of increasing responsibility within many government organizations and continues to advance technology through collaboration with both government and industry to develop next generation training systems. He currently serves as the Government PM and COR for the HP-SON project. Mr. Friedman earned his Bachelor of Science degree in Computer Science in 2000.

AI/ML-driven Network Optimization to enable Synthetic Training and Distributed Simulation

Gregg Patti, Ebony Robinson, Jack Burbank, June Gordon, Todd Lutton, Antonio Fiuza
Sabre Systems, Inc.
Warminster, PA
{gpatti, erobinson, jburbank, jgordon, tlutton, afiuza}@sabresystems.com@email.com

Brad Friedman
U.S. Army Futures Command
Orlando, FL
brad.d.friedman.civ@army.mil

INTRODUCTION

Historically, military training relies on live, in-person training exercises to simulate combat scenarios. While there are definite advantages associated with live training exercises, they also present many disadvantages and challenges. Large monolithic live training exercises require extensive pre-planning, are logistically difficult to execute, are typically expensive to conduct, and are often limited in their scope. To overcome these challenges and limitation to provide an improved training experience, the United States Army developing a ***Synthetic Training Environment*** (STE). This envisioned training architecture combines virtual, constructive, and live training environments into a high-fidelity, network-enabled, collective STE. This federated STE capability will harness the advanced technologies to provide an immersive experience using realistic and accurate representations of actual operational environments and scenarios. These scenarios may include any combination of live combat operations, reconnaissance and surveillance activities, human-machine interaction, civilian population engagements, and other mission activities. This STE capability will allow for a more persistent, continuous, and consistent Warfighter training experience to enhance mission readiness and promote future mission success.

The objective STE vision is technically challenging as it links a heterogeneous, distributed set of Department of Defense (DoD) training systems over a complex network spanning both DoD controlled and managed networks to provide a good user experience. Many of the training elements that will be used in the STE were designed as local, isolated tools and never envisioned distributed interaction with other systems. Another challenge is the increasing complexity and scale of emerging Army systems, such as large Augmented Reality / Virtual Reality (AR/VR) systems, which generates and consumes increasing data volume requiring improved network performance to meet the expected user experience. The STE must also accommodate limitations imposed by bandwidth and computationally constrained environments. The STE will require responsive, low-latency communications networks, which is somewhat analogous to a performance-sensitive distributed game environment. Consequently, methods to establish, maintain and optimize performance across this diverse network and deliver a usable STE training experience are required.

The High-Performance STE Optimization Network (HP-SON) is being developed to address network-related challenges. HP-SON is a holistic architectural framework that utilizes a distributed Artificial Intelligence / Machine Learning (AI/ML)-based approach to dynamically learn networks and user behaviors, and then optimize network performance through a suite of intelligent adaptive methods. HP-SON is an overlay network that consists of network border devices and end-host based client software that performs six core functions:

1. ***Network performance characterization*** – HP-SON characterizes the network performance of inter-connecting components of the STE, including effective throughput, latency, and latency variation.

2. ***User activity characterization*** – HP-SON characterizes how users are utilizing the network and the types and volumes of traffic being offered to the network, including application types, protocols, and traffic volumes.

3. ***Adaptive data compression*** – HP-SON minimizes offered traffic volumes through novel data compression to preserve network capacity and improve performance.

4. ***Intelligent traffic scheduling and forwarding*** – HP-SON provides adaptive differential treatment to different traffic types to better support performance-sensitive applications.

5. ***Dynamic protocol optimization*** – HP-SON dynamically adapts protocol parameters and behaviors based on network characteristics.

6. ***Predictive data caching*** – HP-SON caches commonly accessed data at the network edge to preserve network capacity and improve responsiveness for improved user experience.

These six core functions support the overarching goal to optimize the user experience in challenging network environments. HP-SON cannot make the network better than it is, nor can it exceed theoretical network capacity limits. Rather, HP-SON aims to optimize the user experience by effectively and efficiently using the capacity of the network. Simplistically, HP-SON works to achieve two primary goals:

- ***Saturation Avoidance***- HP-SON attempts to prevent / minimize network saturation conditions (i.e., prevent the network from ever reaching full capacity)

- ***Graceful Degradation*** - As network capacity is reached / approached, HP-SON controls the impact to user traffic.

The remainder of this paper provides an overview of HP-SON and its demonstrated performance to date.

HP-SON OVERVIEW: PREEMPTIVE ADAPTATION THROUGH PREDICTIVE NETWORK ANALYTICS

Modern communications systems and networks commonly employ numerous adaptive techniques, where some behavior is measured, and metrics captured as a set of analytics and then a behavior of the system is adapted based on those analytics. A simple example is Adaptive Modulation and Coding (AMC) where the signal quality of a communication link is measured or estimated and then the modulation and coding utilized on that link is adapted to optimize to that measured or estimated link quality. HP-SON is also predicated on adaptive methods. However, HP-SON aims to opportunistically utilize ***preemptive adaptation***. HP-SON's basic design tenet in this approach is that if you wait until a network change to adapt, you've waited too long, and the user experience will suffer. HP-SON attempts to not only characterize the network and its users, but also to ***learn*** the network and user activity such that network and user activity can be reliably predicted to create ***predictive network analytics*** upon which preemptive adaptation can be conducted. HP-SON's predictive paradigm is depicted in Figure 1.

Additionally, HP-SON aims to leverage the significant advances in the field of AI/ML over the past decade. Indeed, AI/ML-based approaches have been heavily studied over the past decade and there are many applications in the area of communications that are being studied and/or developed. For example, the Third-Generation Partnership Program (3GPP) has established a study group to investigate and identify potential applications for AI/ML for the New Radio (NR) air interface in Release 18 and beyond. There is also a growing base of open literature that discusses potential AI/ML in future 6G cellular specifications (e.g., [1]), with applications typically centered around channel state information (CSI) feedback enhancement, beam management, and positioning accuracy enhancement. However, there is also open literature that discusses the usage of AI/ML algorithms for the purposes of network capacity prediction and planning (e.g., [2]). AI/ML-based algorithmic approaches clearly present many opportunities for performance improvements and optimizations.

In the HP-SON paradigm, the system monitors and builds a time-series history representation of the network performance and network user activity. These time-series histories are used to train an ensemble algorithmic engine comprised of both non-AI/ML and AI/ML algorithms. Once sufficiently trained, HP-SON's algorithmic engine then produces predicted time-series histories of future network performance and future network user activity. These predictive network analytics are then used as the basis for various types of performance preemptive adaptations, including intelligent traffic scheduling, routing and forwarding, protocol optimization, and data caching.

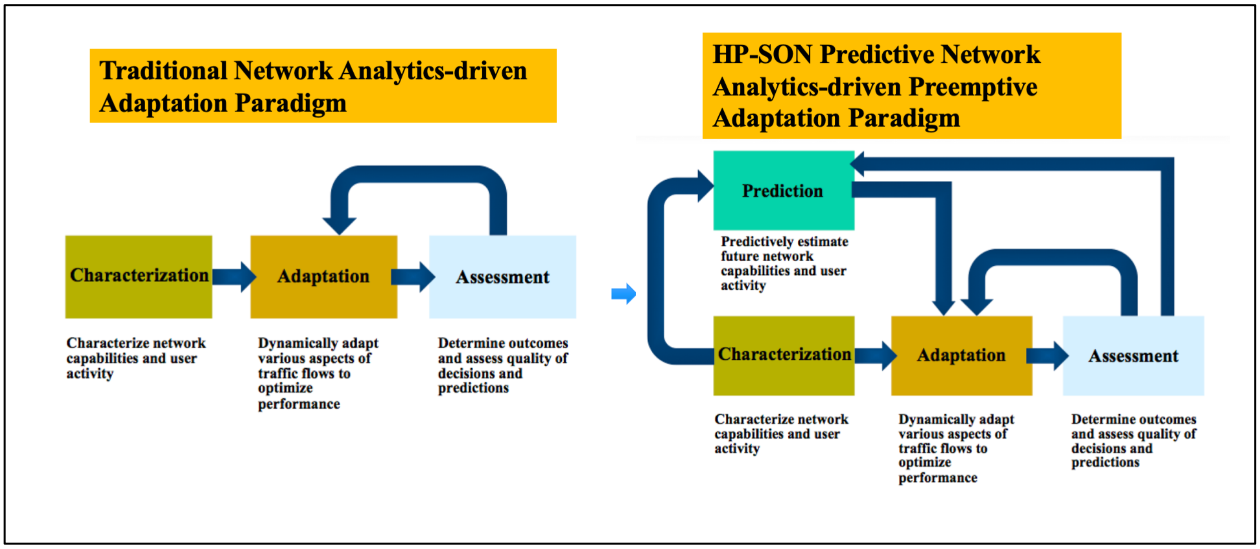

Figure 1. HP-SON's Predictive Paradigm for Network Performance Optimization

HP-SON SYSTEM ARCHITECTURE

HP-SON is an overlay system of distributed components that perform a variety of functions on traffic flows to reduce latency and increase throughput (see Figure 2).

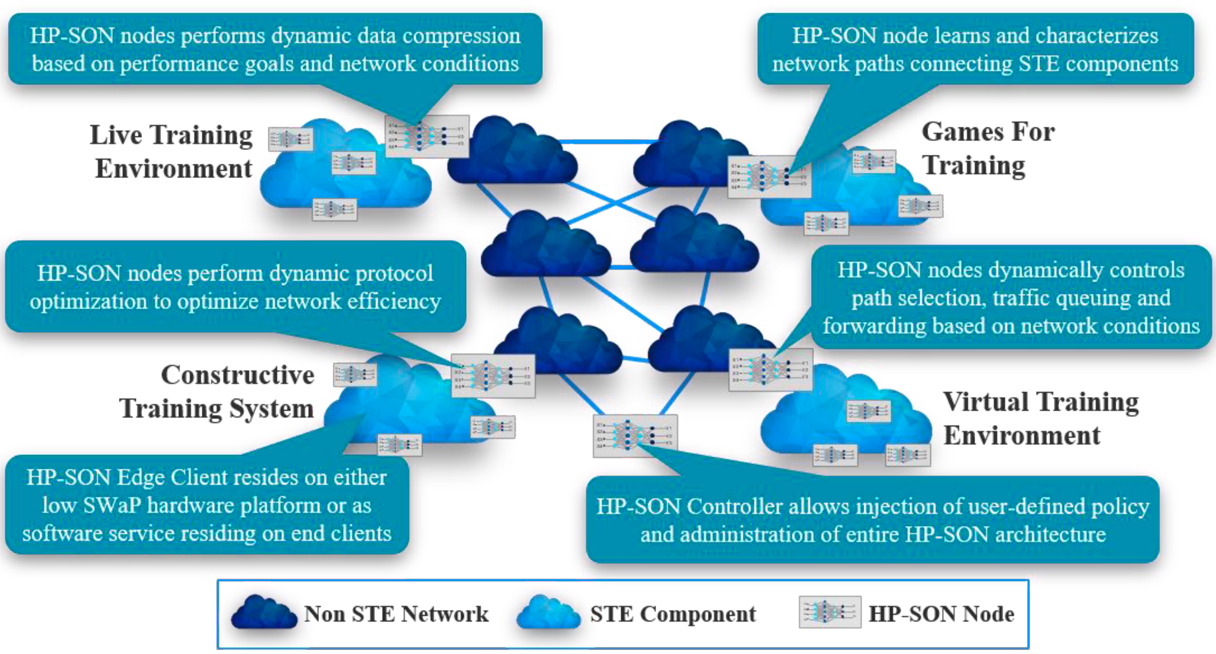

Figure 2. HP-SON System Architecture

The HP-SON overlay network is comprised of border devices and edge-client devices. HP-SON border devices reside on the edge of networks and can be thought of as an AI/ML-enabled smart router and are responsible for network and user activity prediction, traffic scheduling and forwarding, protocol optimization, and data caching. HP-SON client devices are software agents that reside on end-hosts (i.e., user devices) and are responsible primarily for data compression. There is also an HP-SON controller node that is an administrative node used to configure and monitor the overall HP-SON network. These system components are summarized in Table 1.

Table 1. HP-SON Component Summary

Component	Description
HP-Node	• AI/ML-enabled "Smart Router" • Network border device aiming to improve performance over entire network. • Dedicated hardware platform (residing on dedicated commodity hardware). • Supports entire STE end network. • Where network analytics are collected, and predictive analytics are generated. • Acts as a proxy for entire network, providing protocol optimization, and data caching services for the STE end network.
HP-Client	• Existing end-host based device to improve performance for single platform. • Software-based service residing on end-host platform. • Supports individual end-host, providing primarily data compression services for the STE end user.
HP-Controller	• Centralized entity to perform overall system monitoring and control. • Software-based, runs on any commodity hardware. • HP-Node can act as HP-Controller.

NETWORK PREDICTION

Network prediction plays a central role in the HP-SON infrastructure. As previously mentioned, HP-SON builds a time-series history of both network performance and user activity, through both active and passive measurement methods. Network performance parameters that are collected include latency, latency variation, and throughput. Collected user activity parameters include traffic volumes as a function of transport-layer protocol, application, source address, and destination address. These time-series histories are then inputs to HP-SON's algorithmic engine as training data. Once sufficiently trained, HP-SON will then produce predicted time-series history datasets. These predictive analytics are used as the basis for the adaptive methods within HP-SON, including traffic scheduling, forwarding and routing, protocol optimization, and data caching. This clearly places a premium on the establishment of robust and reliable prediction methods.

Historically, many methods have been proposed for the network traffic characterization, including Linear Time Series Models (LTSM), Neural Networks (NNs), Support Vector Machines (SVMs), and Principal Component Analysis (PCA). Additionally, statistical models such as Hidden Markov Models and Bayesian Estimation have been commonly used. These methods can be categorized as either linear or nonlinear prediction methods. One of the most widely accepted models for linear prediction is the Auto-Regressive Integrated Moving Average (ARIMA) family of models. Deep learning models are commonly used for nonlinear prediction. One of the most common types of deep learning approach used for nonlinear prediction are Recurrent Neural Networks (RNNs) or Recurrent Neural Networks with Long Short-Term Memory (LSTM). A good summary of commonly used network prediction methods can be found in references [3]-[5].

Each individual algorithm found in literature has both strengths and weaknesses. Many existing algorithmic approaches are predicated on certain mathematical characteristics or attributes of the underlying datasets under consideration. For example, some algorithmic approaches break down when the underlying datasets have nonlinearities in the data. Many existing datasets are better suited to make predictions on specific timeframes (long term vs. short term). Some algorithms are better suited to identify trends, where others are better suited to identify seasonality in data, or to identify events within the data.

The HP-SON learning engine employs an algorithmic ensemble approach, where the output of multiple independent linear and nonlinear methods are combined to create optimized predictions. This ensemble approach has been crafted such that the overall ensemble does not breakdown under any given statistical scenario in the underlying data and ensures that predictions are accurate across timeframes. HP-SON employs algorithms focused on extracting trend data, seasonality, and event-specific patterns. The result is the ability to reliably predict network traffic and user activity across a wide range of data conditions in a manner that remains robust against underlying changes in the data.

HP-SON Prediction Performance

HP-SON network prediction algorithm performance was evaluated using two datasets containing traffic data (in bits), collected by two different Internet Service Providers. Dataset A was collected by a private Internet Service Provider with centers in 11 European cities. Dataset B was provided by the United Kingdom Education and Research Networking Association. It contains data from the United Kingdom's academic network. Both time series datasets are available through the Time Series Data Library [6]. Prediction performance was evaluated using daily, hourly, and 5-minute interval timeframes.

Prediction algorithms show very promising results using these example datasets. As an example, Figures 3 and 4 show nearly perfect alignment between actual and predicted values, when using actual Internet data from datasets A & B when evaluated in an hourly timeframe. All prediction performance evaluations to date, against numerous datasets have shown similar agreement. Such results are extremely encouraging since proper network prediction enables the remainder of the HP-SON architecture.

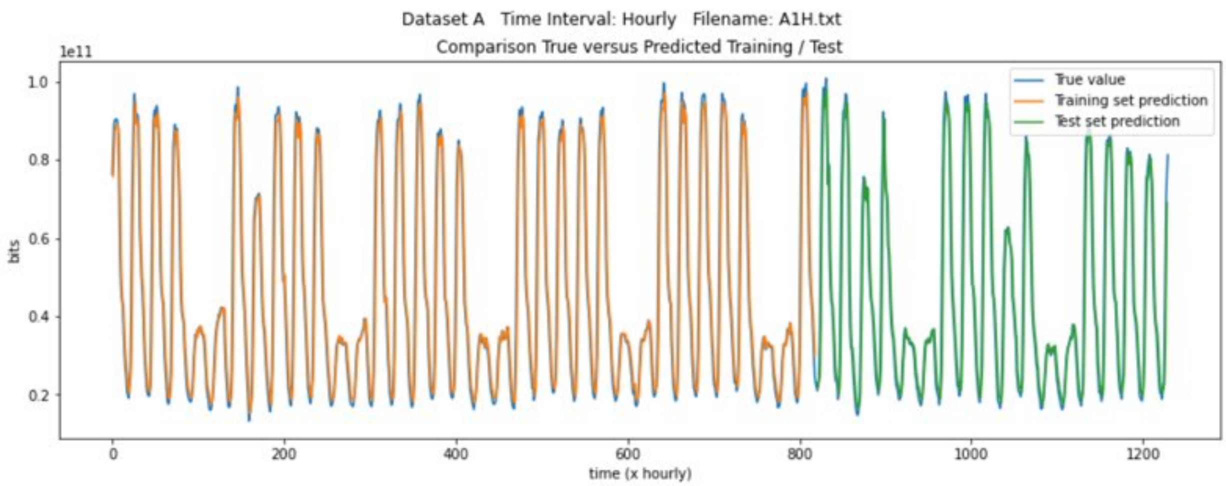

Figure 3. Dataset A values Vs. Algorithm Predicted Values

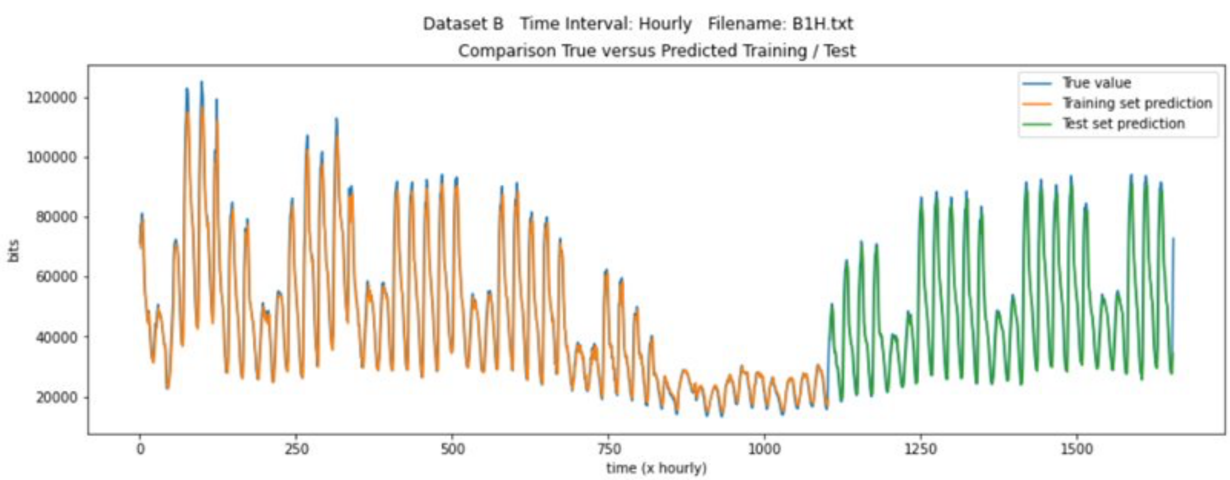

Figure 4. Dataset B values Vs. Algorithm Predicted Values

HP-SON's performance prediction algorithm was quantitatively evaluated using Normalized Root Mean Squared Error (NRMSE) as the evaluation metric, as defined in equation 1.

$$NRMSE = \frac{\sqrt{MSE}}{y_{max} - y_{min}} \qquad (1)$$

where

$$MSE = \sum_{i=1}^{n} \frac{e_i^2}{n} \qquad (2)$$

e is the difference between the true value and the corresponding predicted value
n is the total number of points in the dataset
y_{min} and y_{max} represent the minimum and maximum values in the dataset, respectively.

Table 2 summarizes NRMSE performance for the models in the ensemble algorithmic approach compared with performance of two different singular algorithms (one linear and one nonlinear). Each algorithm achieves strong agreement with the actual data when used individually. However, when utilized in an ensemble approach, even better agreement with real data is achieved.

Table 2. NRMSE Performance of HP-SON Prediction Algorithm Ensemble

NRMSE for Different Models			
Time Series Name	**Linear Algorithm 1**	**Nonlinear Algorithm 1**	**Ensemble**
Daily - 1	.37681746	.30032188	.11558892
Daily - 2	.15712605	.12693225	.08480663
Hourly - 1	.23073049	.08599383	.07601012
Hourly -2	.12690220	.05070741	.04502132
5 min - 1	.19003542	.01799143	.01133993
5 min - 2	.12459774	.00996341	.03185458

Notice how the ensemble approach consistently achieved the lowest NRMSE compared to individual models, apart from the bottom row. These encouraging results solidify the idea of utilizing multiple algorithms in an ensemble approach to effectively predict network traffic and user activity.

DATA COMPRESSION

A key goal of HP-SON is to prevent a network from reaching saturation conditions. HP-SON employs advanced data compression to reduce data volumes offered to the network, thus safeguarding network capacity, and preserving performance across all network users and applications. HP-SON focuses on high-bandwidth applications such as high-resolution imagery and video.

Data compression is a mature field, with several well-established methods widely employed across the Internet for traffic types. Compression methods, such as JPEG and JPEG2000 for images, MP3 for sound, and MPEG 1/2/4 for video, are used in most current network applications. However, there has been promising research over the past several years in the application of ML algorithms to data compression, including use of deep neural networks (DNNs) [7]. HP-SON employs a Deep Neural Network (DNN)-based data compression model. Specifically, HP-SON utilizes a Generative Adversarial Network (GAN) [8] that automatically discovers structure in the input data using an encoder-decoder compression channel. Many proprietary modifications have been made to GAN approaches found in open literature to improve compression performance and fidelity of reconstructed images.

The HP-SON data compression model consists of several components:

- Feature Extractor – discovers structure and reduces redundancy through pyramidal decomposition and inter-scale alignment modules.
- Lossless Coding Scheme – further compresses extracted features by quantizing and encoding through an adaptive arithmetic coding scheme applied on their binary expansions.
- Adaptive Codelength Regularization – modulates expected codelength to target bitrate.

- Reconstruction Loss – penalizes distortions between the target and reconstruction.
- Discriminator Loss – enables visually appealing reconstructions by penalizing discrepancies between their distributions and the targets.

The data compression model has 3 bitrates: low, medium, and high. These various bitrates result in different levels of fidelity (the GAN-based approach is not lossless) but allow the system to adapt bitrates based on predicted network conditions. Figure 5 illustrates the resulting quality from our GAN-based compression. Figure 5 depicts two images side by side. The image on the left is compressed using JPEG and the one on the right is the image after HP-SON GAN-based compression. The PNG-compressed image size is 1.5MB but the HP-SON compressed image size is only 40.5kB after being compressed with a low bitrate. ***This yields a compression ratio of 37.93 over JPEG compression*** and still provides a quality image with little-to-no color or image distortion visible to the naked eye. By utilizing a GAN, even with a low bitrate, compressed images are still of a high-quality and are very difficult to differentiate from the original.

Figure 5. Image using standard JPEG compression (left) and GAN-based (right)

HP-SON data compression algorithms were developed using both the Compress AI [9] and TensorFlow Compression frameworks. These frameworks not only provide the tools for developing new models, but also provide an easy method to compare performance of new models and traditional codecs. Our models were benchmarked against the standard codecs using the Kodak dataset [10]. The Kodak dataset consists of 24 PNG true color images of size 768 × 512 pixels released by the Kodak Corporation for unrestricted research usage. The images are frequently used for generating metrics when comparing traditional codecs with more advanced codecs. HP-SON's GAN-based compression algorithm was compared against JPEG, JPEG2000, and numerous other traditional codecs. The primary metric used to evaluate compression algorithms is Multi-Scale Structural Similarity (MS-SSIM), which is a performance metric specifically designed to match the human visual system. The average MS-SSIM as a function of the bits-per-pixel (BPP) for each compressed image is computed, where BPP is the number of bits needed to represent one pixel. For two images which have been aligned with one another, the luminance, contrast, and structure measures can be given by:

$$l(x,y) = \left(2\mu_x\mu_y + C_1\right)/\left(\mu_x^2 + u_y^2 + C_1\right) \qquad (3)$$

$$c(x,y) = \left(2\sigma_x\sigma_y + C_2\right)/\left(\sigma_x^2 + \sigma_y^2 + C_2\right) \qquad (4)$$

$$s(x,y) = \left(\sigma_{xy} + C_3\right)/\left(\sigma_x\sigma_y + C_3\right) \qquad (5)$$

Then, the Structural Similarity index between the two images can be defined as follows:

$$SSIM(x,y) = [l(x,y)]^{\alpha} + [c(x,y)]^{\beta} + [s(x,y)]^{\gamma} \qquad (6)$$

where α, β, and γ are parameters to define the relative importance of the three parameters.

As an example of the performance of HP-SON's GAN-based compression algorithm, consider Figure 6 which shows MS-SSIM vs. BPP for both HP-SON compression and JPEG compression. It is clear from Figure 6 that the average BPP for a given image quality is significantly less in our GAN-based compression compared to JPEG compression. Similar results have been demonstrated against all standard compression methods. Overall, when using the Kodak dataset as a basis of comparison, our GAN-based compression achieves compression ratios of 20-100 times better than standard PNG-based compression.

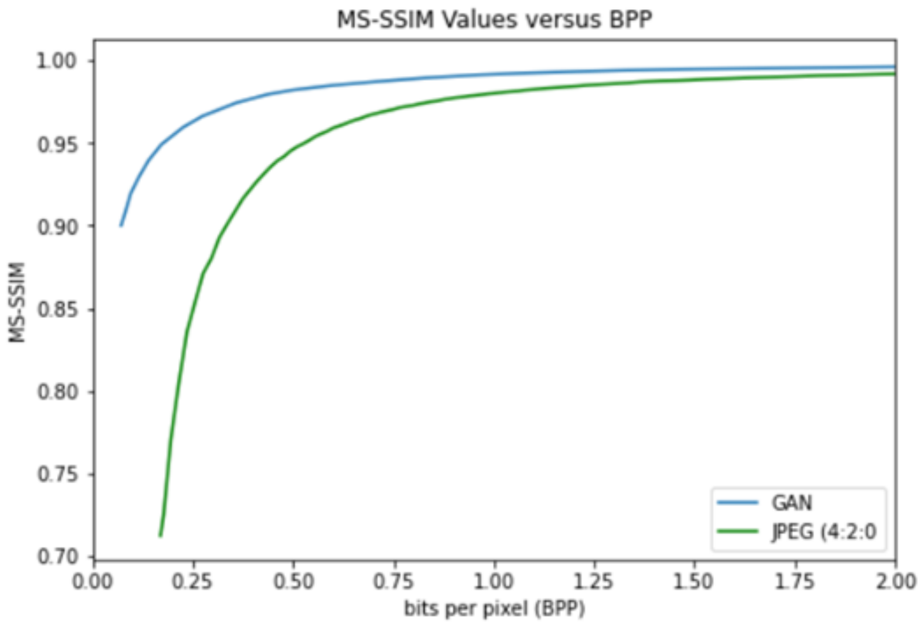

Figure 6. Performance of GAN-based Compression vs. JPEG Compression

INTELLIGENT SCHEDULING AND FORWARDING

Another key goal of HP-SON is to gracefully control the degradation of user experience in cases of network saturation. To accomplish this, HP-SON employs a sophisticated Quality-of-Service (QoS) system of intelligent traffic scheduling, forwarding, and path selection. Here, the goal is that as network performance degrades such that all user traffic cannot be sufficiently serviced, performance-sensitive traffic receives preferential treatment. The intent is to provide different network experiences for each traffic type, based on the needs of that traffic. Effectively, this makes the network appear as a collection of fundamentally different networks to different traffic types. This is done through the creation of virtual network slices and is depicted in Figure 7.

Figure 7. AI/ML-Driven Virtual Network Slicing

This virtual network slicing is driven by the predictive analytics created by the HP-SON network performance and user activity prediction algorithm engine. This capability provides the system with the ability to adaptively control how the network is accessed and utilized to optimize end-to-end network performance, even in times of network saturation or poor performance. HP-SON intelligently chooses which network path STE traffic uses based upon characterization and learned behaviors and performance, with the goal of providing a user experience that best matches the desired performance goals. This intelligent routing is pointedly different from traditional IP network routing approaches. Each HP-Node implements a multi-queue priority forwarding system to forward traffic to available network interfaces to achieve the best overall outcome across all performance goals amongst all applications and users.

The overall goal to make a single network connection appear as various virtual networks, each providing different performance to different types of traffic.

HP-SON utilizes a predictive analytics-driven hybrid of Weighted Round Robin (WRR) and Strict Priority (SP) for scheduling. HP-SON employs WRR, where each queue non-equally takes turns forwarding, to shape ingress traffic, and SP, where certain traffic classes always experience immediate forwarding, which is used for egress priority queuing. HP-SON's forwarding architecture selects the *best* path for outgoing traffic. It utilizes a Software Defined Network (SDN) that separates the control and data planes to allow for fine granularity of packet processing based on user-specified logic. The SDN acts as a switch for priority queuing which sets priority forwarding rules based on the traffic type. While outgoing traffic is going through its path selection, ingress traffic is parallelized and sent back through processing blocks for decompression and protocol processing.

DYNAMIC PROTOCOL OPTIMIZATION

Protocol optimization methods have long been a topic of intense study, and there is a wide array of methods that are available to improve performance. Proxy functions for protocol performance improvements, protocol buffering, protocol aggregation, are examples of classic methods often employed to optimize network performance. However, these methods are typically configured and tuned for specific applications and controlled networks through human-in-the-loop trial and error methods. As mentioned previously, STE components will rely on intermediate network infrastructure to interconnect these assets into the larger STE. Since many of these intermediate transit networks will not be within the direct administrative control of the STE, these manual methods are far less effective.

HP-SON aims to automate this process, and dynamically employ and tune protocol-enhancing methods based on network characterization, previous experiences, and predicted future performance. HP-SON's characterization function helps determine the types of networks that interconnect STE components. This assists in determining what protocol optimizations can be employed at each HP-SON node and how these optimizations are tuned. If low-rate tactical wireless networks are included in the network, perhaps that path should be avoided in preference to other network paths. If subnets include high-delay satellite networks, a performance enhancing proxy (PEP) function should be instantiated. When high-capacity backbone networks are available, fewer optimizations may be required to reduce impact to computational resources. If a path between STE components historically displays high latency variation during certain types of exercises, a traffic smoothing (i.e., throttling) function may be instantiated to mask this variability from the end-users of the STE network.

HP-SON also monitors application traffic trends to make predictive estimates of offered traffic to various STE components during different types of exercises and training activities. This knowledge can be used to proactively instantiate different types of application-based optimizations that can improve network performance. If a small portion of network traffic is expected to have bursty latency-tolerant data transmissions during an upcoming time interval, network slices could be proactively reconfigured to provide more resources to accommodate constant bit rate (CBR) applications such as voice. If upcoming data is expected to be dominated by real-time streaming video, then network slices could be proactively reconfigured to optimize latency for that traffic. There is a wide range of possibilities when we can predictively and autonomously optimize the network (i.e., the virtualized HP-SON network).

PREDICTIVE DATA CACHING

HP-SON will provide a predictive data caching capability that is based on a distributed ML decision engine. This decision engine learns user application trends and then predicts future data needs. This engine offers HP-SON the ability to conduct application traffic analysis—to monitor, understand, and anticipate trends in application traffic. With this HP-SON will be able to forward deploy commonly accessed data types within the network to create edge data stores. This provides the opportunity to reduce end-to-end latency, reduce data rate requirements, improve reliability, and improve the overall user experience. This will, in turn, improve STE responsiveness. HP-SON will leverage knowledge of future periods of low utilization to perform these automated data fetches and forward placement.

SUMMARY

Synthetic training environments represent a fundamental paradigm shift, and the STE aims to fundamentally change the way in which our armed forces are trained. However, there are difficult challenges in achieving the STE vision. The networks that inter-connect the geographically dispersed components of the STE must provide reliable, low latency communications to support this complex online game. HP-SON's suite of AI/ML algorithms aim to mitigate the network challenge and provide the reliable, low latency connectivity required to fully realize the STE vision. HP-SON aims to enable a smooth delivery of AR/VR graphics and data in the STE, resulting in a good user experience including in locations with reduced network capabilities. HP-SON will never improve a network past its theoretical capabilities. Instead, it makes the best of the current network situation, and optimizes the network for the best user experience.

The STE will grow and change over time, resulting in different hardware, network architectures, traffic patterns, and user activity trends. However, HP-SON has been developed in a highly modular manner, allowing it to morph and change to adapt to the changing STE network. This makes HP-SON a highly scalable, long-term solution to the Army STE network.

REFERENCES

[1] Y. Shi, et. al, "Machine Learning for Large-Scale Optimization in 6G Wireless Networks," 10.48550/arXiv.2301.03377, January 2023.

[2] Y. Li, et. al, "Analytics and Machine Learning Powered Wireless Network Optimization and Planning," 10.48550/arXiv.2209.06352, September 2022.

[3] G. P. Zhang, Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model, Neurocomputing 50, pp 159-175, 2003.

[4] R. Madan and P. S. Mangipudi, Predicting Computer Network Traffic: A Time Series Forecasting Approach using DWT, ARIMA and RNN, Proceedings of 2018 Eleventh International Conference on Contemporary Computing (IC3), 2-4 August 2018, Noida, India.

[5] P. Cortez, M. Rio, M. Rocha and P. Sousa, Multi-scale Internet Traffic Forecasting using Neural Networks and Time Series Methods, Expert Systems, May 2012, Vol. 29, No. 2, pp 143-155, Wiley-Blackwell.

[6] R. J. Hyndman, Time Series Data Library. [Online] https://github.com/FinYang/tsdl

[7] S. Ma, X. Zhang, C. Jia, Z. Zhao, S. Wang, and S. Wang, Image and Video Compression with Neural Networks: A Review, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 30, No. 6, June 2020.

[8] O. Rippel and L. Bourdev, Real-time Adaptive Image Compression, Proceedings of the 34th International Conference on Machine Learning, PMLR 70:2922-2930, 2017.

[9] J. Bégaint, F. Racapé, S. Feltman, and A. Pushparaja, CompressAI: A PyTorch Library and Evaluation Platform for End-to-End Compression Research, ISO/IEC JTC1/SC29/WG11 MPEG2020/m54467, July 2020.

[10] Kodak Corporation True Color Images. [Online] http://r0k.us/graphics/kodak/