This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

**1. Presentation Title:**
Emulation Modeling for Development of Cyber-Defense Capabilities for Satellite Systems

**2. Presenter:**
Robert G. Cole, Ph.D., R&D S&E, Experimental Cyber Initiatives Department, Sandia National Laboratories, rcole@sandia.gov, +1.443.910.4420

**3. Short Biography of Presenter:**
Ph.D. in Theoretical Chemistry Iowa State University, Postdoctoral Positions in Theoretical Chemistry at Yale and Brown Universities. Distinguished Member of Technical Staff at Bell Laboratories, Research Faculty Department of Computer Science at Johns Hopkins University and Research staff at the Applied Physics Laboratory, Research staff at Army CERDEC and Army Research Laboratory, prior to joining Sandia National Laboratories in 2015. At Sandia, working on applications of Artificial Intelligence to cyber defense of Industrial Control Systems.

**4. Other Authors:**
J. Fustos, R&D S&E, Sandia National Laboratories, jfustos@sandia.gov
B. Hart, R&D S&E, Sandia National Laboratories, bhart@sandia.gov
B. Hill, R&D S&E, Sandia National Laboratories, brehill@sandia.gov
S. Wade, R&D S&E, Sandia National Laboratories, swade@sandia.gov
A. Cooper, Graduate Student, Computer Science, North Carolina A&T State University, acoope@sandia.gov
D. Cardona, Graduate Student, Statistics, Purdue University, cardonad@purdue.edu
A. Sabbaghi, Associate Professor, Applied Statistics, Purdue University, sabbaghi@purdue.edu
C. Bullard, Founder/CEO, QoSient, LLC, carter@qosient.com

**5. Point-of-Contact if other than Presenter:**
NA

**6. Keywords:**
Satellite Systems, Emulation Modeling, Cyber Defense, Generative Methods, Machine Learning

**7. Has this abstract been cleared by your organization?**
Yes

**8. Abstract:**
The objective of this project was to develop a novel capability to generate synthetic data sets for the purpose of training Machine Learning (ML) algorithms for the detection of malicious activities on satellite systems.  The approach experimented with was to a) generate sparse data sets using emulation modeling and b) enlarge the sparse data using Generative Adversarial Networks (GANs).  We based our emulation modeling on the Open Source NASA Operational Simulator for Small Satellites (NOS3) developed by the Katherine Johnson Independent Verification and Validation (IV&V) program [1] in West Virginia. Significant new capabilities on NOS3 had to be developed for our data set generation needs.  To expand these data sets for the purpose of training ML, we experimented with a) Extreme Learning Machines (ELMs) [2] and b) Wasserstein-GANs (WGAN-GP) [3].

Training, testing, and validating ML algorithms for the purpose of detection of malicious activities requires extremely large volumes of data.  These data are typically not available from deployed satellite systems.  This is because the

missions of these deployed satellites are related to National Security, and the data are consequently tightly held. Furthermore, for research purposes, and the training of future researchers, it is useful to have access to non-classified data sets for experimentation purposes.

Sandia National Laboratories (Sandia) has a long history in the development of cyber mitigation systems for National Security missions and has often relied upon emulation modeling, Hardware-in-the-loop (HITL) and analysis tools for this purpose. Within Sandia, these are collectively referred to as Emulytics™ [4] capabilities. Within the domain of networked applications, Sandia’s Emulytics™ capabilities include Minimega [5] and Firewheel [6] for Enterprise Technology (ET) environments, and SCEPTRE [7] for Operational Technology (OT) Industrial Control Systems (ICS) environments. However, for satellite system environments, no such Emulytics™ capability existed. The purpose of this work was to experiment with the development of an Emulytics™ capability at Sandia for the experimentation, design, and training of ML cyber-defense systems for satellite systems.

This was an ambitious objective as previous Sandia Emulytics™ programs were in development for typically five to ten years or more and continue their developments even to this day. We leveraged the NOS3 emulation system to jump start our efforts. However, this system was designed primarily for system test of pre-deployed satellites in development, not for the purpose of generating realistic operational data. So, in this presentation, we will discuss how we converted a manual simulator into an automated end-to-end satellite simulation system. Our simulator uses innovative approaches to automatically plan, execute, and evaluate satellite imaging missions similar to a real ground crew and satellite. The creation of this exemplar imaging mission was inspired by the MultiThermal Imaging (MTI) satellite system [8] managed by Sandia.

Our modifications to the original NOS3 system allowed us to capture realistic ground station/satellite traffic and automatically label it for future machine learning use. Further capabilities were developed to improve overall realism of the data sets and to improve quality of data collection. These included a HITL capability, incorporation and modification of the Argus flow collector [9] on the ground station to satellite link, serial bus collectors and modifiers, and integration of data extraction into Elasticsearch for access to downstream data analysis tools. With these modifications for improved data set creation and collection, we were able to experiment with ELMs [2] and WGAN-GPs [3] for data set synthetic expansion and to demonstrate improvements in resulting ML algorithms for cyber-defense purposes. Finally, in the process of performing these experiments, we realized that the GANs-WP combined with the capabilities of the ARGUS traffic flow monitor [9], potentially form a new anomaly detection capability for cyber-defense applications. More studies are required to strengthen this later proposition.

References:

[1] “NASA IV&V” - https://nasa.gov/centers/ivv/jstar/nos3.html

[2] Huang, G.-B. and D.H. Wang,”Extreme learning machine: a survey”, Int. J. Mach. Learn. & Cyber., vol.2, p.107-122 (2011).

[3] Gulrajani, I., et al., “Improved Training of Wasserstein GANs”, (https://arxiv.org/abs/1704.00028), (2017).

[4] “Emulytics™” - https://sandia.gov/emulytics/

[5] “minimega” - https://sandia.gov/emulytics/uur_minimega-fact-sheet.pdf/

[6] “Firewheel” - https://sandia.gov/emulytics/uur_firewheel-fact-sheet.pdf/

[7] “SCEPTRE” - https://osti.gov/servlets/purl/1376989/

[8] “MTI Satellite” - https://earth.esa.int/web/eoportal/satellite-missions/m/mti/

[9] “Argus” - https://openargus.org/

**9. External communication statement:**

*Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525.*