2021

Automated Cyber Operations Data Mission Replay

Prince, Charles D.; Singh, Gurminder; Shaffer, Alan B.

Monterey, California: Naval Postgraduate School

https://hdl.handle.net/10945/69772

Final Report
Automated Cyber Operations Mission Data Replay

This research is supported by funding from the Naval Postgraduate School (NPS), Naval Research Program (NRP) (PE 0605853N/2098). NRP Project ID: NPS-21-M006-B
Topic Sponsor: Marine Corps Forces, Cyberspace Command (MARFORCYBER), RDTE
Sponsor Representatives: Capt Victor Vega, USMC; Capt Danial Salazar, USMC

NPS Group Members
Principal Investigator Mr. Charles Prince
Principal Investigator Dr. Gurminder Singh
Technical Researcher Dr. Alan Shaffer
NPS Student Capt Mark Petersen, USMC
Resultant thesis: https://calhoun.nps.edu/handle/10945/68370

Executive Summary

The research explored the capability of replaying cyber mission data (CMD) within the Persistent Cyber Training Environment (PCTE) infrastructure through the development of an application. The primary research question of how network specifications can be extracted from log data to create a digital twin network (DTN) of an operational network was answered by a proof-of-concept tool called the Automated Cyber Operations Mission Data Replay (ACOMDR), which can input intrusion detection system (IDS) Zeek CMD from the Big Data Platform-Cyber Hunt & Analytics Operation System (BDP-CHAOS) and produce a network specification to create a network similar to the original network (Petersen, 2021). A secondary research question examined how transformation of log data into executed scripts on a DTN can be performed, and was answered by the method ACOMDR working with Puppet. The ACOMDR to Puppet interface and Application Protocol Interface (API), which have yet to be created, will be used to replay events and to control nodes, as well as interface with the PCTE automation tool. A completed working system will also require additional Puppet functionality to control the nodes. Another secondary research question investigating desirable runtime environment features to control execution of cyber mission scripts, must still be determined.

Recommendations:

- Complete the development of a working prototype by completing the ACOMDR to Puppet interface and API to replay events, and to better extract network specifications from Zeek Data.
- To create a higher fidelity digital twin network, refining the collection process that creates the Zeek data would also help by capturing more usable network data.
- To create networks faster, one should be able to use ready-made networks, and to correlate incoming Zeek data to best fit a set of ready-made networks.
- Investigate incorporation into PCTE a much higher fidelity virtualization system like Xen/Drakvuf, a malware investigation tool.

Summary Project Background

The purpose of the study was to develop an application for replaying CMD within the PCTE infrastructure. It required creation of a virtual digital twin network based on network classification data. The Zeek classification data was accessed through BDP-CHAOS/PCTE and then used to create a raw database of unique datatypes, classified and placed in another database to be used in the creation of the virtualized digital twin network. The Zeek data was not rich in terms of classifiable information, and it is possible that some valuable data was purposefully filtered out since it may have come from a classified network.

Introduction

One of the foundations of winning battles is maintaining comprehensive situational awareness of the battlespace. Cyberspace is a uniquely challenging battlespace in that cyberspace operations (CO) are conducted in abstract digital environments, which introduce new complexities to developing and maintaining situational awareness. This research has begun the development of new capabilities to enhance visualization of CO by instantiating a digital twin of an operational network for replaying the events that transpired during a mission. Since PCTE is still under development the possibility of near-term incorporation of these capabilities into the operational PCTE is high. Creating a digital twin of an operational network and replaying real world CMD would support operator training as well as developing and testing new strategies for Offensive Cyber Operations (OCO) and Defensive Cyber Operations (DCO).

Study Benefits

Joint Cyberspace Operations Forces should benefit from this study as the ACOMDR proof-of-concept tool may lead to timesaving automation of creating scenarios based on real world operational networks. By saving time and increasing fidelity in the creation of CO scenarios, the cycle from incident to training can be shortened, and PCTE is envisioned to be used by all commands.

Study Process

The study reviewed previous cyberspace training solutions to develop a proof-of-concept tool to extract network specifications from CMD and to set conditions to replay CMD on a digital twin network. The focus of this study was to analyze CMD logs provided by Marine Corps Forces Cyberspace Command (MARFORCYBER) and develop a software environment to process and extract network specifications to instantiate a digital twin network on the PCTE.

The project was divided into two phases. The first phase consisted of researching, collecting data on, and analyzing CO training environments and CMD logs to extract network configuration data to be used to instantiate a digital twin network within PCTE. The second phase consisted of developing software to process an operational network's CMD logs, and extract specific details from the log data to construct a network specification. The tool built in this study, ACOMDR, allows for future development and integration with both existing PCTE capabilities and external data repositories.

Background

This study found the current state of the art in cyberspace training environments offers a range of capabilities and resources to replicate real world scenarios, heighten situational awareness of the battlefield, and increase training and readiness of personnel through high-fidelity digital twin networks and systems. Key findings from these technologies emphasize the high degree of portability and API of the Java software platform, the value of graphical depictions of network actions for increased situational awareness, and consolidated training resources under a single host system. PCTE has the capacity to offer all of these capabilities in a cohesive interoperable environment. Replaying CMD on a simulated operational network requires the establishment of the desired network in a virtual environment. Currently, PCTE offers automated solutions to build and configure virtual networks. There are a variety of interface methods with PCTE's networking tool that include a variety of scripting languages and API tools to process CMD and interface with PCTE.

Methodology and Status

The ACOMDR Application must have incoming data through the PCTE framework accessing BDP-CHAOS data. The application then parses the data and stores it in a database that categorizes it as data that can be used to create a Digital Twin Network (DTN), or data used to create and run events on the DTN systems. The ACOMDR application consists of three modules and a SQL database running on a Java platform.

The ACOMDR application is designed to be hosted from within the PCTE infrastructure and allow users to process collected CMD within the PCTE environment. The ACOMDR application will enable connections to external data repositories, which enables the PCTE to process collected mission data for user visualization, training, and future mission rehearsals. The figure below (figure 1) shows the ACOMDR tool, labeled as Cyber MDR Application, and its event relationships, dark red, with PCTE, along with its external event relationships, labeled light red, to the database and external intrusion detection/prevention system.

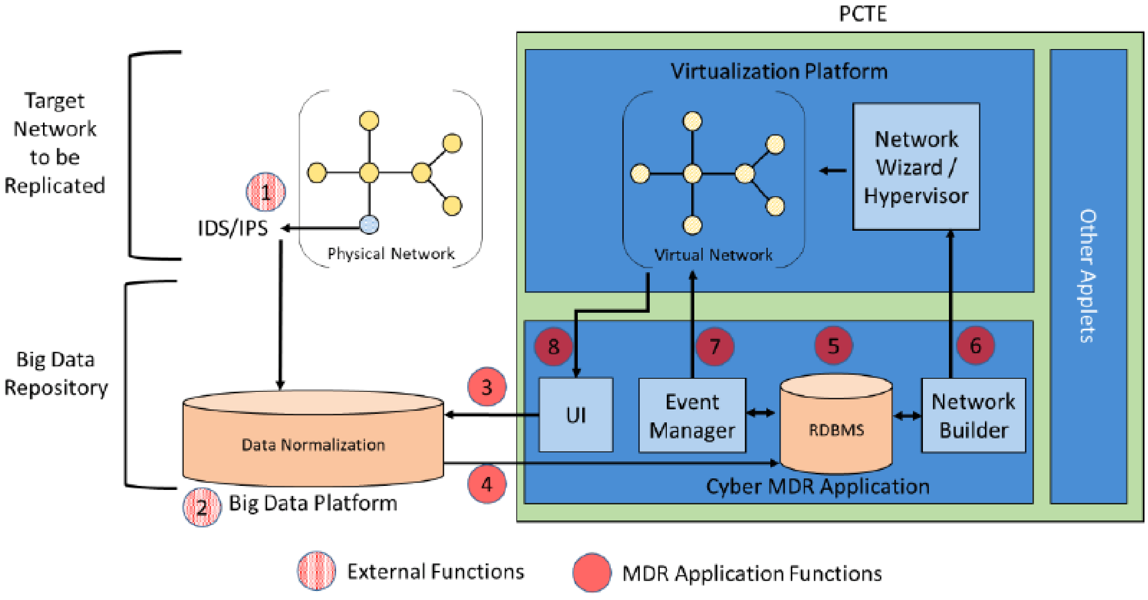

Figure 1

The numbered external and MDR application functions show the chronology of events for ACOMDR to acquire data and perform network building and event management. Data is created by the IDS (1) and stored in (2) BDP-CHAOS. ACOMDR (3) pulls data from BDP-CHAOS and stores it (4) in its SQL database. The data is then either used by (5) the Network Builder to create a network of systems (6), and or by the Event Manager (5) to create events (7) in PCTE.

The ACOMDR proof of concept application has been partially completed; further work is needed to complete its scoped-out functionality, including instrumentation and interface to control remote systems in the DTN to replay events. Bolt, a tool within the Puppet open-source software project, has been identified as a potential tool to provide this functionality. Puppet is already a part of the PCTE architecture, so integration of Bolt is not expected to be challenging. Work needs to be performed to create an interface between PCTE, ACOMDR, and Bolt, and instrumentation needs to be created to run the events. Additionally, ACOMDR needs to be instrumented to create the DTN, and interface calls need to be created to work with PCTE to create the DTN.

The existing ACOMDR prototype tool can be more effective in creating a DTN if the process it uses to collect the data is improved. For example, a mechanism can be created to refine the Zeek log data extracted by the IDS from events and network data. In addition, the IDS data collected on the original network could be modified to extract fine grain networking and events that will lead to better DTNs and events orchestrated by ACOMDR.

Unusual Difficulties Encountered

This study occurred during the height of COVID19, which created delays in connecting and interfacing (networking) with the groups in charge of working with the jointly operated PCTE tools and environments.

Findings

We found that network specifications can be generated from IDS BDP-Chaos Zeek data, and a proof-of-concept tool, ACOMDR, was created to perform this task. We believe that a working system can be created to ingest IDS Zeek logs and use the data to create a representative network and control nodes on that network to simulate IDS incidents, but a working prototype for the system has yet to be created.

We recommend that this research effort be continued to develop a fully finished prototype of ACOMDR. Completing the ACOMDR to Puppet interface and API to replay events and to control nodes, as well as the interface with the PCTE automation tool, will be essential to completing the additional functionality to control the network nodes for attack replay. Additionally, the ability to search the BDP-CHAOS databases will result in increased functionality for the ACOMDR application. Creation of a graphical user interface to display views of DTN topology and cyber events over time would also enhance system capability. More work on inferring original network specifications based on Zeek data and translating it to the YAML specification needs to be performed. To support this, improving the Zeek data from the IDS will greatly enhance the tool's capabilities.

We also believe that adding pre-configured networks to PCTE and then finding the best representative network according to incoming IDS Zeek data would be a valuable timesaving effort for PCTE. A ready-made-network would be a network that is already created and ready to use. In addition, we believe that a study to add a Xen/Drakvuf high-fidelity virtualized servers would be highly valuable to the study and training of malware for PCTE.

Recommendations

Completion of the development of a working prototype by completing the Automated Cyber Operations Mission Data Replay (ACOMDR) to Puppet interface and application protocol interface (API) to replay events and control nodes would require additional work to develop the interface and API. Additional work on refining the extraction of network specification from the existing Zeek data to gather better network specifications to refine the process needs to be performed. This project would complete the existing study.

An investigation into creating readymade networks would require modifying the ACOMDR to be able to use this new feature and test the concept of correlating and analyzing algorithms used on incoming Zeek data to best fit ready-made networks. The hypothesis of this project would be that by using ready-made networks time could be saved by end users waiting for a network to be created.

An investigation into generating higher fidelity Zeek data by creating a better intrusion detection system/Big Data Platform-Cyber Hunt & Analytics Operation System such that the Zeek data can be extracted to make higher fidelity digital twin network specifications. This process of intrusion detection system extraction would have to be looked at and analyzed. This project's hypothesis is that a higher fidelity digital twin network and event data could be created.

An investigation of the incorporation into Persistent Cyber Training Environment (PCTE) of a much higher fidelity virtualization system (HFVS) like Xen/Drakvuf, would involve looking into PCTE to see what functions could be connected into an HFVS and may require development of infrastructure to support the connection. There may be additional functionality that must be created in PCTE as well as in the HFVS, as well as adaptation of ACOMDR into HFVS. The hypothesis for this project is that a network could be created that could test, and develop malware that is aware of virtualization.

References

Petersen, M. (2021). *Automated cyber operations mission data replay* [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun, https://calhoun.nps.edu/handle/10945/68370.