FALL 2022 – COS 397 COMPUTER SCIENCE CAPSTONE PROJECT PROPOSALS

Project Title: Abnormal Network Traffic Flow Dashboard Tool

By: Ben Drozdenko, Ph.D.

Brief Description (approximately 500 words)

You and your team work for a company that has an internal network enclave that is not connected to the Internet. To protect the confidentiality and integrity of the data on this enclave, you will need to develop a system that imports and visualizes network flows to alert presence of malicious network activity by zero-day exploits and insider threats. Since a commercial off-the-shelf (COTS) intrusion detection/prevention system (IDS/IPS) is attuned to typical network traffic for internet-connected servers, these systems have proven to be ineffective in the past, and so a customized in-house solution is needed. However, the company does not have sufficient revenue to afford a full-time network analyst, and can only afford partial attention by a local area network (LAN) Administrator when suspect activity is occurring. For this reason, the company's chief technical officer (CTO) has asked you to develop a Dashboard Tool with an Artificial Intelligence / Machine Learning (AI/ML)-based solution that monitors real-time network traffic flows, identifies abnormal traffic flows, and notifies the LAN Admin immediately when consistent, sustained abnormal activity has been observed. A sample pipeline that shows the flow of data in blue and automated processes in black from the network hardware to the LAN Admin is shown in Fig 1.

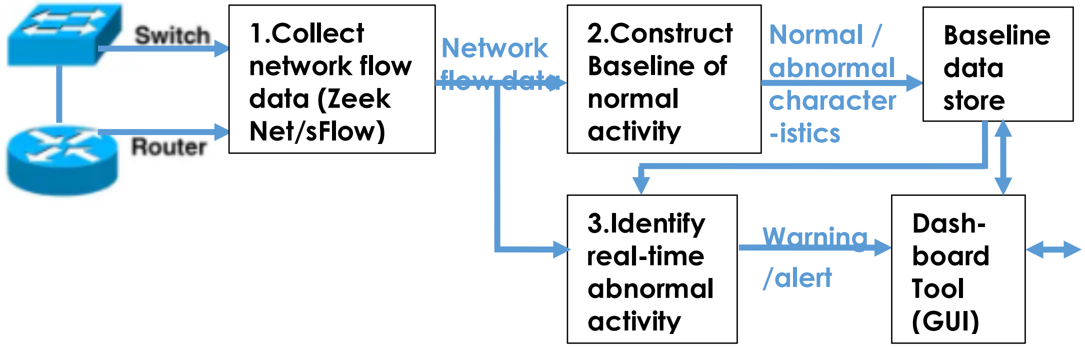
Switch
Router
1.Collect network flow data (Zeek Net/sFlow)
Network flow data
2.Construct Baseline of normal activity
Normal / abnormal character -istics
Baseline data store
3.Identify real-time abnormal activity
Warning /alert
Dash- board Tool (GUI)

Figure 1: Abnormal Network Traffic Flow Dashboard Tool

The CTO has made it clear that you should use network flow metadata, such as NetFlow or sFlow, which are collected at the switches and routers to accommodate large amounts of traffic and real-time concerns. Flow (a.k.a. dataflow) monitoring supports vital network management tasks such as traffic engineering, anomaly detection, detecting worms, scans, and botnet activities, and forensic analysis. Flow-based network monitoring relies on collecting information about packet flows (a sequence of related packets) as they traverse routers, switches, load balancers, network visibility switches, and other devices. Network elements that support traffic monitoring protocols such as NetFlow and sFlow extract critical details from packet flows, like source, destination, byte and packet counts, and other attributes. This "metadata" is then streamed to "flow collectors" so that it can be stored and analyzed to produce network-wide views of bandwidth usage, identify abnormal traffic patterns that could represent possible security threats, and zero in on congestion sources. Flow-based monitoring provides significant advantages over other network monitoring methods. It can capture substantially more detail about network traffic, and it is also significantly less expensive and easier to deploy than raw packet capture. Furthermore, since flow generation capability is now a built-in feature of almost all modern network equipment, it provides a pervasive monitoring footprint across the entire network. Alternately, the CTO has been researching Zeek, an open source network security monitoring tool (formerly Bro), as a software-based alternative to collecting network flow data, but he is unsure how the format of Zeek logs vary from NetFlow and sFlow data. For more information on Zeek, see https://zeek.org/.

Goals for the project (approximately 50 words)
The goal is to develop a Dashboard Tool with an Artificial Intelligence / Machine Learning (AI/ML)-based solution that monitors real-time network traffic flows, identifies abnormal traffic flows, and notifies the LAN Admin immediately when consistent, sustained abnormal activity has been observed.

Total Duration / Elapsed Time [in weeks]: 24 weeks

External Schedules / Deadlines (conferences paper deadline? Meetings?) [if any]: Plan to have weekly check-in meetings (e.g., Friday mornings) to report on the progress the team has made on the capstone project and giving the team an opportunity to ask for design specifications or technical feedback.

Learning Objectives for student teams:

This project will give the student teams an opportunity to learn about the way that network traffic can be monitored and stored by intrusion detection systems (IDSs) and intrustion prevention systems (IPSs). In addition, it is an opportunity to learn about the Artificial Intelligence / Machine Learning (AI/ML) techniques used by these systems, including but not limited to supervised learning (e.g., Random Forest), deep neural networks, self-supervised learning, and transfer learning.

Expected Project Experiences (select from the list):

Problem definition
Project scope definition
Design and implementation of research methodology
Use of applied statistics
Data analysis
Workflow analysis
Development of functional specifications
Identification of and negotiation for needed project resources
Examination of an unfamiliar technical area
Identification of others' technical expertise
Identification and evaluation of alternatives
Development and presentation of recommendations
Responsibility and accountability for a discrete product
Role definition in a task group and participation in group dynamics

Recommended experience (What operating system is required? What programming language? Other skills?):

I'd recommend students who have experience with Unix operating systems such as Red Hat Enterprise Linux (RHEL) 8.5, but deployment on a Windows 10 OS is also acceptable. The Python programming language would work best, but the students are welcome to write the programs in the language of their choice. Experience with the Anaconda and Jupyter Notebook environments for Python programming is a plus.

Expected Outputs/Products and likely requirements (specific programming language, programming framework, operating system, integration with existing software, web-based requirements, *etc.*):

The project will require software development, and so the working Dashboard will be the main product. The other deliverables shall include the following:

- Project report hardcopy and electronic copy on DVD-R , consisting of:
    1. Design documentation including use-cases, requirements, class diagrams, system level diagrams, test plan, and test results as appropriate.
    2. User documentation which may include setup, usage, API, etc
    3. Source code for all developed components

Past experiences by the client (If software already exists, what is wrong? What has worked in previous versions, and what has not?):

The client has already tried deploying commercial IDSs and IPSs to tactical systems, but these have had no success due to the unique nature of the network traffic on their closed enclaves. The client has collected network traffic data (which cannot be shared unfortunately), and so is interested in ideas for automating its analysis and producing a set of necessary and sufficient relevant data to be consumed by LAN admins.

Proposed Testing Plan (How will the team test their product? Do you have recommended/required testing strategies? What resources are available (test platform, stand-alone network, *etc.*)? Is test data available?):

The team should plan to develop their own test routines to perform verification and validation of the dashboard tool. For test data, the team may want to observe the characteristics of malicious network traffic from public sources such as the University of New South Wales (UNSW) dataset available online here: https://research.unsw.edu.au/projects/unsw-nb15-dataset. Establishing a baseline of flow features of both normal and malicious traffic would improve the accuracy of AI/ML techniques such as clustering (e.g., KMeans).

Benefits to U Maine:

New contacts with Naval Undersea Warfare Center Division Newport (NUWCDIVNPT) means that more U Maine students will know about employment opportunities available for scientists and engineers at our beautiful Newport, RI facility overlooking the Narragansett Bay. Increased knowledge, skills, and abilities in Computer Science and related fields of Data Science and Cybersecurity can be applied to undersea tactical platforms (e.g., submarines) to produce cyber-resilient systems that protect the national interest.

Project Sponsor(s):

Ben Drozdenko, Ph.D.
Cyber Science & Technology (S&T) Technical Lead
Undersea Warfare (USW) Combat Systems Department, Cybersecurity Division (Code 255)
401-832-3992 / benjamin.m.drozdenko.civ@us.navy.mil
Naval Undersea Warfare Center Division Newport (NUWCDIVNPT)

Other Resource People:

Aiden Lammert, NUWCDIVNPT ISSO Team (Code 2554)
401-832-4016 / aiden.r.lammert.civ@us.navy.mil

Software/server access required:

The Project team is expected to use their own server for hosting the Dashboard for their development efforts. Use of software tools such as Anaconda and Jupyter Notebook for developing Python code is recommended.