26 June, 2014

MEMORANDUM FOR: AFIT/ENG

ATTENTION: MAJ DUBE

FROM: Capt Phillip Warner (GCE-15M)

SUBJECT: Thesis Prospectus:  *Automatic Configuration of Embedded Device SCADA Honeypots*

1. The goal of this research effort is to develop a proof of concept architecture for automatically configuring supervisory control and data acquisition (SCADA) honeypots. SCADA devices, which control much of the world’s critical infrastructures, have been shown to often be highly vulnerable yet exposed to the Internet. Many efforts have been undertaken to develop decoys, or honeypots, of these devices in order to characterize attacks or aid law enforcement in attributing attacks against SCADA networks. Unfortunately, since SCADA devices typically are proprietary and unique, one emulation solution for a particular vendor’s model will not likely work on other devices. This research will provide a scalable solution to that problem by developing an architecture to automatically configure honeypot emulation of arbitrary SCADA devices. Emulation will be performed by replaying saved data response streams according to the interrogation of the honeypot by another machine. These response streams will be generated by capturing and parsing traffic between a real live SCADA device and an interrogator. Each of the honeypots will be implemented on Raspberry Pi (RPi), and a laptop will serve as a master to control the various slave emulators.

2. This research is a follow-on project to an AFIT thesis report from June 2014 titled *Toward Automating Web Protocol Configuration For a Programmable Logic Controller Emulator* (Fink, 2014). The former research effort was primarily focused on automatically emulating a device’s web interface by downloading its web pages and serving them via a python web server as requested. This research, however, will use an algorithm similar to what was used in a 2012 AFIT thesis report titled *External Verification of SCADA System Embedded Controller Firmware* (McMinn, 2012) in order to parse captured traffic. Rather than using the captured traffic to check firmware against, however, this research will use it to build emulation profiles. Additionally, this research will capture a wider variety of conversations and use different protocols and devices than the previous mentioned works. For the proof of concept architecture Allen Bradley Control Logix devices will be used, and the CPU module and Web cards will be emulated over EthernetIP. Captured traffic to emulate will include firmware download, firmware upload, and live run responses. The interrogators used to generate this traffic will be primarily the vendor’s device programmer (e.g. RS Logix 5000 for Allen-Bradley devices), but will also include common scanners such as the Nessus SCADA scanning profile. If time permits, additional devices from other vendors and/or additional protocols will be emulated.

3. The automatic honeypot configuration research will be done in several phases. During the first phase familiarization with the PLCs, programmers, protocols, and typical device usage must be done. Traffic will be inspected, parsed, and saved using Wireshark and tcpdump. At least two packet captures must be done for every conversation between the interrogator (e.g. the device programmer or Nessus) and each of three target PLCs. Two packet captures are required so stateful portions of packets can be identified. Additionally, three firmware versions for each target PLC will be tested. Testing multiple firmware versions is particularly important for the firmware upload and download conversations.

The second phase will be to automatically create emulation profiles (collections of responses to given inputs). A script will be developed to parse the saved traffic from the previous phase and identify sections to emulate. This will be saved in a hash table of responses using the requests as a key.

Once the profiles have been created for each PLC and corresponding firmware version, the third phase will be to test the profiles and how they are selected. Profile selection will be done by:

1. a command from the master to emulate a particular model PLC and firmware version
2. using secondary information sent to the RPi emulator
3. a command from the master to emulate a device at a particular IP address
4. emulating devices based on the packets sent to the emulator

Profile selection method 2 will use information such as a profile imported from the master or a Nessus XML report. Method 3 will cause the RPi emulator to scan a device to generate secondary info for method 2. Method 4 is fully automatic emulation. All RPi honeypots will use basic out-of-band logging implemented on a shared RPi log server with an attached USB hard disk.

Resources required to complete the research effort include a dedicated non-edu network PC for PLC traffic analysis and code development, a virtual machine with RS Logix 5000 installed, target PLCs (Allen-Bradley ControlLogix 5555, 5561, and 5571) for emulation, a USB hard disk, and at least two RPi devices with dual network connectivity. Additionally, a hub will be needed to ease capturing traffic between the PLCs and interrogator laptop, and a router will be needed to setup a private network for the master laptop, RPi emulators, and RPi log server. Finally, for any additional testing involving equipment from other vendors (e.g. Siemens) corresponding PLCs and programmer software will be required.

4. The output of this research effort will be an architecture using multiple RPi devices and embedded software that will automatically build emulation profiles, select a profile to emulate based on the criteria described above (paragraph 3), and emulate (within the scope of this research) the functionality of the selected target device. The effectiveness of the emulation will be measured in two ways. One, the interrogator’s reaction to the honeypots will be compared to that of the PLCs to see if the scanning tools perceive them to be the same. For instance, if RS Logix 5000 can be used to successfully “program” the ladder logic for a honeypot, then the test is a success. If a Nessus scan results in identical fingerprints for the honeypot and real PLC then the test is a success. The second measure of effectiveness will be done by comparing captured responses byte for byte from the honeypots and the real PLCs. The goal is that the responses would be similar enough to fool knowledgeable adversaries into thinking they are interacting with the actual device. Finally, further testing would need to be done to ensure that correct profiles are selected based upon the criteria described in paragraph 3. For testing the fully automatic emulation (method 4 above) this would be done by sending the emulator randomized packets with known corresponding profiles. The number of correctly selected profiles will be recorded. The speed at which the emulator can handle profile changes could also be tested by sending the randomized packets at increasingly faster speeds.

5. The results of this research could be a great benefit for any community attempting to make use of SCADA honeypots as it would provide a foundation for arbitrary PLCs to be emulated rapidly versus weeks or months of effort. These honeypots would be useful for academic research and cyber exercises such as in AFIT’s Mobile Device Security and Cyber Defense courses. The NSA might also use them for training and exercise purposes. Finally, being able to automatically configure SCADA honeypots would allow incident response teams to rapidly deploy honeypots to locations with suspected security breaches. The honeypots could then be used to log malicious activity and catch the offenders.

6. Proposed thesis committee:

a. Dr. Barry Mullins, Chair / Thesis advisor \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ *(signature)*

b. Mr. Juan Lopez, Committee member \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ *(signature)*

c. LTCOL Mason Rice, Committee member \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ *(signature)*

7. Sponsor: Department of Homeland Security ICS-CERT

a. Name: Neil Hershfield

b. Title: DHS ICS-CERT Technical Lead

c. Organization: ICS-CERT

d. Section or Division: NPPD/CS&C/NCSD/US-CERT

e. Grant Number: HSHQDC-11-X-00089

f. Telephone number: 1-877-776-7585

g. E-mail address: [ics-cert@dhs.gov](mailto:ics-cert@dhs.gov)

8. The coursework completed at AFIT has been planned to provide the foundation of knowledge required to successfully complete this thesis. Courses of particular applicability to the research topic include:

* CSCE 560 Introduction to Computer Networks
* CSCE 629 Cyber Attack

Additionally, familiarization of Allen-Bradley devices, ICS protocols, and the RS Logix 5000 software will be attained through self-study.

Phillip Warner, Capt, USAF

GCE-15M

1st Ind, AFIT/ENG

MEMORANDUM FOR AFIT/ENG

I approve/disapprove the above thesis prospectus and thesis committee. This prospectus will be maintained in the student’s file. The thesis should be prepared in accordance with the AFIT Thesis Guide. Good luck!

THOMAS DUBE, Maj, USAF

Chief, Computer Science Division

Department of Electrical and Computer Engineering