Firmware Indicator Translation: Cyber Threat Intelligence Big Data Analytics Proof of Concepts Using ELK and STIX™

Christian O. Hunt, *Principal Security Engineer and Lead Researcher, New Context*, Kelly Cullinane, *Director of Utility and Energy Services, New Context*, Andrew Storms, *VP of Product and Security Services, New Context*

***Abstract*-- In partnership with Idaho National Laboratory, New Context has completed a multi-year project to examine how to incorporate cyber threat intelligence using ELK and STIX™ into big data analytics as part of INLs Firmware Indicator Translation (FIT) project. New Context researchers were able to prove that open source tools combined with an open, unified standard can have a significant impact on the monitoring of cyber threat intelligence.**

I. Introduction

New Context is a services and product company that protects data and its movement in highly regulated industries. Since January of 2014, the company has been a strategic partner in the California Energy Systems for the 21st Century (CES-21) project working directly with Pacific Gas and Electric, Southern California Edison, San Diego Gas and Electric, Idaho National Laboratory, and Lawrence Livermore National Laboratory. The company has expertise in designing and building secure and compliant big data platforms in critical infrastructure industries such as energy, financial services, industrial IoT, and transportation.

Further, New Context is heavily involved in creating standards for cyber threat intelligence frameworks. The company is advancing the state of the art in cybersecurity through open standards such as Structured Threat Information Expression (STIX™), Trusted Automated eXchange of Indicator Information (TAXII™), and Open Command and Control (OpenC2).

Recently, New Context built a proof of concept that utilized ELK to store binary data from malicious files. New Context then used STIX™ patterning as a vendor-agnostic, repeatable method to describe and search for malicious content in these files.

Given that STIX™ is the desired indicator and observables format for encapsulating data about firmware and identifying unwanted code, New Context was primed to take a research role for the Firmware Indicator Translation (FIT) project.

II. Technical Gap

In the area of cyber threat intelligence, indicators based on a deep understanding of compiled executables usually take a back seat to data artifacts such as log messages, configurations, file hashes, or network flow data. Binary patterns are an incredibly useful tool to identify threats in executables and other contexts not normally considered.

The FIT project focused specific attention on searching numerous indicators to find matches in firmware and then translating them into popular threat feeds while having little to no impact on general operations.

In addition, the FIT project aimed to leverage the indicator work done at Idaho National Laboratory along with ongoing automated response projects – specifically the software developed for Structured Threat Intelligence Graph (STIG) which uses the STIX™ standardized data structures. The language created for STIX™ indicators of compromise (IOC) based on cyber observables and remediation actions has been very effective in identifying and responding to cyber events in substation automation test configurations.

III. Project Objectives

The objective of FIT was to sift through the volumes of indicators to find the applicability to firmware, feed firmware indicators into general-use threat feeds, and apply indicators at the firmware level without impacting operations.

In order to do this, New Context researchers aimed to execute the following two projects to demonstrate these objectives:

- Demonstrate how to search binary data in ELK using STIX™ patterns
- Demonstrate ELK searches for binary analysis data storage and search

Each project is described within the document and demonstrates the challenge, our approach, and the results.

IV. Strategy

In order to execute the objectives for FIT, New Context researchers leveraged the following tool and formats.

A. STIX™

Structured Threat Information Expression (STIX™) is an open-standard format used to exchange cyber threat intelligence (CTI) in a machine-readable format.

B. JSON

JavaScript Object Notation (JSON) is an open-standard

format that uses human-readable text to transmit data objects consisting of attribute-value pairs and array data types.

C. ELK

ELK is a schema-less database and set of management and processing tools -- ElasticSearch, Logstash, and Kibana -- used for building scalable distributed databases.

V. PROJECTS

A. Project 1: Using STIX™ Patterns to Search Binary Data in ELK

1) Approach:

New Context created an ELK instance and leveraged example observable and indicator data from the Idaho National Laboratory FIT project in order to demonstrate the ability to quickly search a set of observables for identified Indicators of Compromise (IOC).

In order to introduce randomness and diversity of data, the team created a tool to generate random binary observables defined using the STIX™ payload_bin artifact object. These observables were submitted to the Elasticsearch database along with periodically including binary data that would match the pre-defined indicators. This allowed a stream of observables to be simulated. In order to have the Elasticsearch database properly index the payload_bin artifact, it was necessary to convert the artifact data from Base-64-encoded binary data and represent it as an escaped hexadecimal string.

Once a collection of observables had been submitted to Elasticsearch, New Context used a tool to convert the STIX™ pattern to an Elasticsearch query to "find" any observable that matched the supplied pattern.

This operation of searching the binary data within ELK using a STIX™ pattern proved to be fast and repeatable. Furthermore, by utilizing the STIX™ standard the use case shows that STIX™ can be utilized across many different vendor and technology stacks with little or no modification.

2) Challenges:

New Context researchers encountered two main challenges when executing this project:

- How to effectively use STIX™ patterns to search a database of binary observables.
- How to translate a vendor-agnostic STIX™ pattern into a vendor-specific format.

3) Results:

After completing the project, New Context researchers identified the following key results:

- The proof of concept was able to prove that binary cyber observables (such as firmware) could be stored in an Elasticsearch database.
- The international standard of STIX™ patterns can be programmatically converted to ELK queries and retrieved at an acceptable speed for big data analytics.

B. Project 2: ELK for Binary Analysis Data Storage and Search

1) Approach:

Most people believe cyber observables are typically data artifacts such as log messages, configurations, hashes or network flow data. However, cyber observables may also be the outcome of pre-processing events. By storing the results of a pre-analysis, such as the output of compiler agnostic function detection in binaries, the observables become more contextual to the problem.

The approach used the output from Nucleus, a tool that attempts to identify functions and subroutines in disassembled code, and constructing observables. These observables consisted of code fragments, the callees and callers of that code fragments, and disassembly of the fragment.

This allowed for queries to be performed against the output of the disassembly findings, for example, specific addresses, function callees or callers. Then, any aspect of the observable can be queried to potentially construct call/flow graphs from the data or apply more advanced patterns looking for multiple code fragments with specific characteristics.

2) Challenges:

New Context researchers encountered two main challenges in executing this project:

- How to effectively store binary analysis data in an Elasticsearch database so it becomes searchable content similar to cyber observables.
- How to effectively construct queries against these data points and structures in order to identify potentially unwanted firmware or firmware components.
- STIX™ by default requires binary data to be encoded using Base64 and ELK will not index base64 binary blobs by default without use of 3rd party plugins.

3) Results:

After completing the project, New Context researchers identified the following key results:

- While most approaches choose to store cyber observables directly in a database for search, the project demonstrates how preprocessed analytics results may also be stored.
- The approach demonstrates how a prior knowledge of a system may be used to create a more effective and efficient search.

VI. CONCLUSION

The project showed that open-source tools combined with open standards can be utilized for cyber threat intelligence monitoring and response in an effective manner leading to increased cybersecurity resilience and shorter mean duration to response. In addition, the project recommended several modifications and extensions to the STIX™ standard to help facilitate the handling of these types of data structures.

For example, a possible change to STIX™ would include moving common data structures that would normally be encapsulated into an artifact object into first-class STIX™ objects where appropriate. For example, if a list of subroutine offsets is commonly used, such as a list of debug symbols,

creating a STIX™ object to represent these data points in a sane manner would be of benefit.

VII. Appendix I

A. About New Context

Since January of 2014, New Context Services, Inc. has been engaged with California utilities, Idaho National Labs, and Lawrence Livermore National Labs to assist in the California Energy Systems for the 21st Century project initiated by the California Public Utilities Commission. Throughout this project, New Context has acted as a subject matter expert to the partner utilities and engaged national labs.

New Context is considered the foremost authority in extending STIX™ to support the needs of the electrical industry. Historical submissions to extend the standard for electrical utility uses have included:

- Author of the STIX™ patterning quick reference guide
- Created utility-specific STIX™ extensions for DNP3 and ModBus
- Developed tools to represent ICS specific temporal event indicators of compromise
- Enabled mechanisms to perform multi-sensor correlation for OT networks

New Context protects data and the movement of data in highly regulated industries. Our Lean Security methodology integrates security into software development, critical infrastructure, and architecture. The company has expertise and experience in designing and building secure and compliant big data platforms in critical infrastructure industries such as financial services, industrial IoT, and transportation.

B. New Context Consultant Skills

Our cybersecurity consulting services focus on helping our customers evolve their capabilities across the entire spectrum of people, process, and technology. Included below are a few example descriptions of our team's skills and experience.

1) Principal Security Engineer / Subject Matter Expert:

a) Description:

The Principal Security Engineer is a subject matter expert. This person drives content and thought leadership generating innovative approaches to address security problems. They provide technical leadership with a hands-on approach. This person demonstrates strong skills to architect, design, implement, maintain and operate information system security controls and countermeasures.

b) Similar Titles:

Principal Analyst, Principal Engineer, Principal Architect, Principal Security Architect

c) Experience & Education:

Minimum of 20 years experience and Bachelors Degree

2) Principal Software Engineer / Subject Matter Expert:

a) Description:

The Principal Software Engineer is a senior expert with extensive knowledge in a designated field or discipline. Provides insight and advice concerning task or project strategic direction and outcomes. May contribute to the evaluation, analysis, and development of recommended solutions. Resolves complex problems, which require an in-depth knowledge of the subject matter.

b) Similar Titles:

Principal Analyst, Principal Engineer, Principal Architect, Principal Functional Specialist, Principal Application Engineer

c) Experience & Education:

Minimum of 20 years experience and Bachelors Degree

3) Senior Program Manager:

a) Description:

Provides oversight and executive level management to overall contract operations. Responsible for managing multiple deliverables and workstreams. The Program Manager maintains and manages relationships at the senior level with customers, partners and other 3rd party organizations. They ensure quality standards and work performance, organizes and oversees work efforts, assign resources, manages personnel, provides risk management.

b) Similar Titles:

Senior Strategy Associate, Executive Strategy Associate

c) Experience & Education:

Minimum of 15 years experience and Bachelors Degree

C. New Context Industry Experience

1) Open Standards For Cyber Threat Intelligence:

New Context is advancing state-of-the-art cybersecurity through open standards such as Structured Threat Information Expression (STIX™), Trusted Automated eXchange of Indicator Information (TAXII™), and Open Command and Control (OpenC2). We are a leader in developing global standards in Cyber Threat Intelligence (CTI) and building infrastructure around referenceable architectures. New Context works closely with OASIS and the most innovative standards organizations to advance transparent standards framework development.

In addition to pioneering the STIX™ patterning language, New Context has led many initiatives within STIX™ to specifically represent electrical utility needs such as DNP3 and Modbus. Historical accomplishments in STIX™ have included representation of complex ICS indicators of compromise, extensions for electrical utility communication protocols, temporal events, and multi-sensor correlation. New Context is also the author of the popular STIX™ Patterning language quick reference guide.

2) Industry Thought Leadership:

New Context is focused on supporting industry and our government to solve the complex security issues facing our critical infrastructure. To that end, when asked, we provide our thoughts and guidance to those tasked with protecting our energy infrastructure. Below is a recent example of congressional testimony before the U.S. Senate. We've provided a link to the full testimony and an excerpt of our CEO's remarks before the committee.

https://www.c-span.org/video/?436330-1/hearing-focuses-energy-infrastructure-cybersecurity

"For the past three years we have been working closely with

Southern California Edison, Pacific Gas and Electric, and San Diego Gas and Electric, in partnership with Idaho National Lab and Lawrence Livermore National Lab, to assist in advanced cyber-security research for machine-to-machine threat detection and response within the energy industry. This project is referred to as California Energy Systems for the 21st Century. That work has resulted in our involvement in the STIX™/TAXII™ and OpenC2 standards that are becoming the default for governmental agencies, enterprises, and information sharing communities (ISAOs & ISACs) to distribute cyber-threat intelligence rapidly. There are five areas of advanced cyber-defense that I will be discussing in my testimony: Identity, Trusted Data, Attributed Isolated Networks, Threat Detection & Sharing, Automated Response and Remediation…"

VIII. APPENDIX II

Table 1
Project Terminology

Term	Definition
Cyber Observable	STIX™ Cyber Observables document the facts concerning what happened on a network or host, but not necessarily the who or when, and never the why. For example, information about a file that existed, a process that was observed running, or that network traffic occurred between two IPs can all be captured as Cyber Observable data.[1]
CTI	Cyber Threat Intelligence
DNP3	DNP3 (Distributed Network Protocol) is a set of communications protocols used between components in process automation systems.[2]
ELK	The ELK Stack is a collection of three open-source products — Elasticsearch, Logstash, and Kibana — all developed, managed and maintained by Elastic.[3]
Executable	Causes a computer "to perform indicated tasks according to encoded instructions", as opposed to a data file that must be parsed by a program to be meaningful.
FIT	Firmware Indicator Translation
Indicator	Contains a pattern that can be used to detect suspicious or malicious cyber activity.[4]
Kibana	Kibana is an open source data visualization plugin for Elasticsearch.[5]
Logstash	Logstash is an open source, server-side data processing pipeline that ingests data from a multitude of sources simultaneously."[6]
Modbus	Modbus is a serial communications protocol originally published by Modicon (now Schneider Electric) in 1979 for use with its programmable logic controllers(PLCs).[7]
Nucleus	Compiler-Agnostic Function Detection in Binaries[8]
payload_bin	A STIX™ artifact[9]
STIX™	Structured Threat Information Expression (STIX™)[10]
STIX™ pattern	STIX™ patterning language to enable the detection of possibly malicious activity on networks and endpoints[11]
TAXII™	Trusted Automated Exchange of Intelligence Information (TAXII™)[12]
TC	Technical Committee
Oasis	OASIS is a nonprofit consortium that drives the development, convergence and

[1] http://docs.oasis-open.org/cti/stix/v2.0/stix-v2.0-part4-cyber-observable-objects.html
[2] https://en.wikipedia.org/wiki/DNP3
[3] https://www.elastic.co/elk-stack
[4] https://docs.google.com/document/d/1IvkLxg_tCnICsatu2lyxKmWmh1gY2h8HUNssKIE-UIA/_edit#heading=h.muftrcpnf89v
[5] https://www.elastic.co/products/kibana
[6] https://www.elastic.co/products/logstash
[7] https://en.wikipedia.org/wiki/Modbus
[8] https://bitbucket.org/vusec/nucleus
[9] http://docs.oasis-open.org/cti/stix/v2.0/stix-v2.0-part5-stix-patterning.html
[10] https://oasis-open.github.io/cti-documentation/stix/intro.html
[11] http://docs.oasis-open.org/cti/stix/v2.0/stix-v2.0-part5-stix-patterning.html
[12] https://oasis-open.github.io/cti-documentation/

	adoption of open standards for the global information society.[13]
OpenC2	The OpenC2 fulfill the needs of cybersecurity command and control in a standardized manner.[14]

[13] https://www.oasis-open.org/org

[14] https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=openc2