

# Fixing Codified Cyber Analysis

August 2022

William Raymond Brant





#### DISCLAIMER

This information was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness, of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. References herein to any specific commercial product, process, or service by trade name, trade mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

## **Fixing Codified Cyber Analysis**

**William Raymond Brant** 

August 2022

Idaho National Laboratory Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517, DE-AC07-05ID14517

Will Brant | University of Wyoming | Mentors: Rita Foster, Zachary Priest | Technical Leads: Bryan Beckman, Jeremy Jones



## **Overview**

- Many organizations collect and codify cyber threat data for further analysis.
- However, a lack of consistency renders much of the data ineffective.

## **What Can Change**

- JSON (Java Script Object Notation) is the data codification standard(1).
- Threat analysis can be converted from one JSON
   schema to another with the use of interpreters.
- Standard enforcement of JSON can create meaningful data for all use cases.

### **Schemas**

- . STIX (Structured Threat Information Expression)(2)
- 2. CSAF (Common Security Advisory Format)(3)
- 3. Synapse Hyper-Graph Query Data(4)

Bottom: STIX ShadowPad Campaign Model

Right: : CSAF Example (<a href="https://json.csaf.io/">https://json.csaf.io/</a>) (graph is static and non-interactive)

Top: Synapse Data (query output is like JSON but fails JSON validation)

# Projects

#### STIX Enforcer

- Goal: Enforce a specific sub-schema on STIX JSON for standardized bundles. Standardized bundles can then be fed to a machine learning model.
- Work Accomplished: Developed four fixers to write updated JSON to a separate file ensured the logging of meaningful data.
- Notes: Subschema verification can be implemented easily without major conversions.

#### CISA(Cybersecurity and Infrastructure Security Agency)-Scraper

- Goal: Homogenously extract data from CISA reports (HTML), convert CISA data into CSAF and STIX, and convert between CSAF and STIX without data loss.
- Work Accomplished: Refactored existing scraping code and ensured that CSAF and STIX data could be converted without data loss.
- Notes: While conversion between JSON schemas is possible, the process takes a considerable amount of time to program. The
  relationship between the schemas is not one-to-one—STIX bundles are comprised of a list of objects and CSAF documents are comprised
  of nested dictionaries.

#### CSAF-Validator

- Goal: Automate the validation of a large number of CSAF documents.
- Work Accomplished: Implemented multi-processing, integrated automation with a server launched within github workflow.
  - Notes: The CSAF validation testing files did not catch every error. Schema validation is computable, but the ensuring of meaningful cyber-security data is not.

#### Synapse-to-JSON

- Goal: Convert Synapse Storm queries to JSON and Golang structs, certifying standardization.
- · Work Accomplished: Scraped synapse queries, pulling out and storing relevant data.
- Notes: Converting JSON-like data to JSON does not fall within regular languages; thus, regular expressions alone cannot change the
  queries into proper JSON. However, the conversion is context free, which is why some elements can be converted via regular expressions.



## Conclusion

- Codifying cyber-threat data into JSON does not imply that the data is meaningful
- If every organization makes a JSON schema, no schema will be useful.
- The work accomplished in the STIX Enforcer, CISA-Scraper, CSAF-Validator, and Synapse-to-JSON projects enforces homogenous cyber threat data.
- This consistent data provides the opportunity for truly effective analysis.

www.inl.gov INL/MIS-22-68201

