INL/MIS-20-59767-Revision-0

Russians in the Grid Enrichment Process

October 2020

Bryan R Beckman

Changing the World's Energy Future

Idaho National Laboratory

INL is a U.S. Department of Energy National Laboratory operated by Battelle Energy Alliance, LLC

INL/MIS-20-59767-Revision-0

Russians in the Grid Enrichment Process

Bryan R Beckman

October 2020

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

“Russians in the Grid” Enrichment Process

Bryan Beckman, Senior Infrastructure Protection
Idaho National Laboratory – Idaho Falls, ID USA
National and Homeland Security, Critical Infrastructure Protection
bryan.beckman@inl.gov

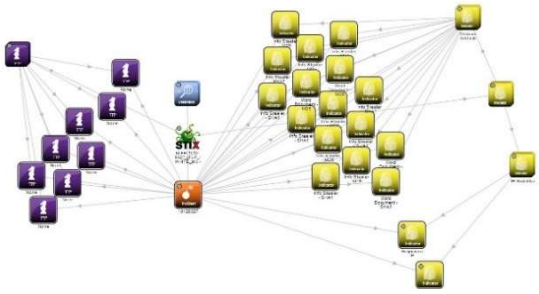
Abstract - Threat intelligence sources are numerous, and the quality of their intelligence is equally variable. The need for quality, actionable, and sharable intelligence is becoming increasingly important as cyber-threats continue to increase. In this paper, we describe the process we used to enrich U.S. CERT’s Technical Alert entitled “Russian Government Cyber-Activity Targeting Energy and Other Critical Infrastructure Sectors,” affectionately referred to as “Russians In the Grid,” from a series of separate but related reports to a final, actionable, unified Structured Threat Information Expression (STIX) v2 representation of the report. In the process, we detail the tools and procedures we used, some of which are automated and others highly manual. We identify some of the limitations of current STIX enrichment tools and techniques, and propose enhancements to allow for more effective enrichment and better overall threat intelligence to be used for more effective threat data-sharing, as well as machine learning-based predictive threat work being conducted at Idaho National Laboratory.

A few key items need to exist for threat intelligence to be useful and effective. The threat indicators and courses of action must be stored in a format allowing it to be easily sharable and actionable. If threat intelligence is generated and stored in a proprietary format, that intel automatically has reduced value due to its limited audience, whereas intel generated and stored in standardized format, such as Structured Threat Information Expression version 2 (STIX2) [1], affords the ability for anyone with the desire or need to be able to read, understand, and utilize the data contained inside. Properly utilizing and supporting a standardized format such as STIX2 allows for threat intelligence to be distributed and shared amongst agencies, corporations, and asset owners. The effective sharing of intel makes it increasingly difficult for a threat actor to utilize the same attack on multiple targets.

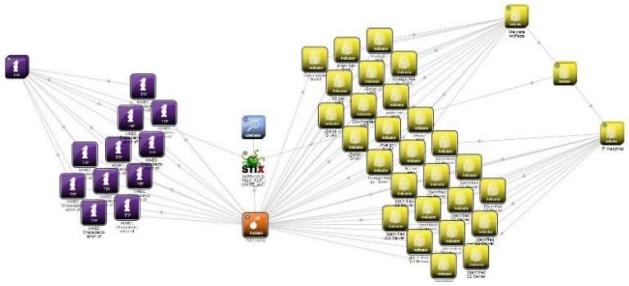
Threat intelligence that is not actionable and unable to tell a complete story also has limited value. Knowing the static hashes for a particular strain of malware, though useful, is not as beneficial as knowing the complete story of an attack, including the sequence of events and the generalized attack patterns and tactics that an adversary is currently using. Actionable threat intelligence can also include recommended courses of action to take in the event that the malicious activity described by the intel is detected again.

On March 16, 2018, the U.S. Department of Homeland Security (DHS) and the Federal Bureau of Investigation (FBI) published a joint Technical Alert (TA18-074A) entitled “Russian Government Cyber-Activity Targeting Energy and Other Critical Infrastructure Sectors” [2], which details cyber-activities by the Russian government as early as March 2016. The TA also included seven STIX1 extensible markup language (XML) files, as well as associated portal document files (PDFs) and a single comma-separated values (CSV) file. Through a manual multistep process, we were able to take the STIX1 files and the textual data from the remainder of the report and ultimately generate a single STIX2 bundle, which provides a more complete picture of the activity described in TA18-074A. This enriched and more complete picture of the activity includes a total of 126 STIX2 objects with 133 associated relationships.

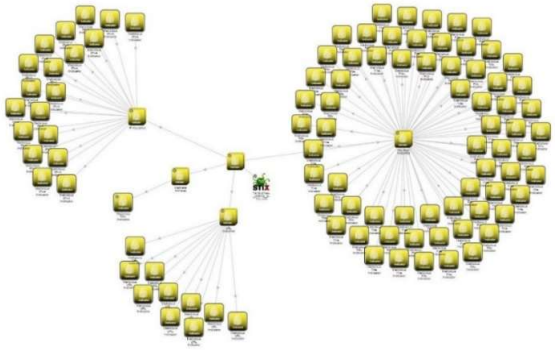
Several tools and techniques were used to enhance the STIX output of the TA. STIX-viz [3] was used to graphically view the STIX1 XML files originally provided with the report. Figure 1 shows all seven of the included STIX1 files as graphs. In these graphs, the yellow icons represent indicator objects, purple icons denote tactics, techniques, and procedures (TTP) objects, orange icons signify incident objects, and red icons show malware objects, as seen in the legend at the bottom of Figure 1.



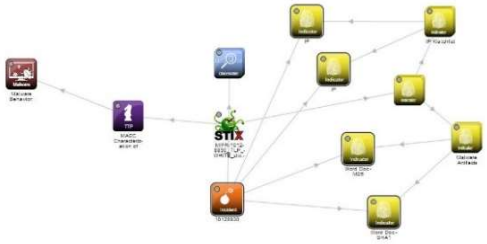
Malicious Tools (MIFR-10128327)



Malicious Tools (MIFR-10127623)



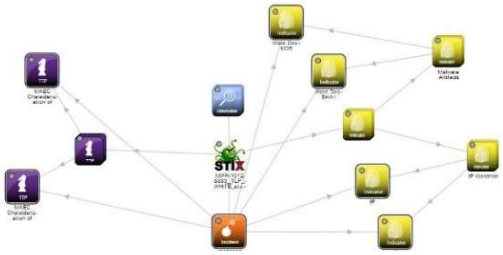
Indicators (TA18-074A)



Malicious Documents (MIFR-10128830)

Malicious Documents (MIFR-10128336)

Malicious Documents (MIFR-10135300)



Malicious Documents (MIFR-10128883)

STIX-viz Legend

Figure 1. Seven original STIX1 files viewed in STIX-viz.

Each of the seven STIX1 files were processed using OASIS' STIX2-Elevator [4], a sample of which is provided in Figure 2. During the elevation process, some data is lost as not all entries properly convert from STIX1 to STIX2. An attempt will be made in a subsequent step to recover/replace this lost data through manual enrichment.

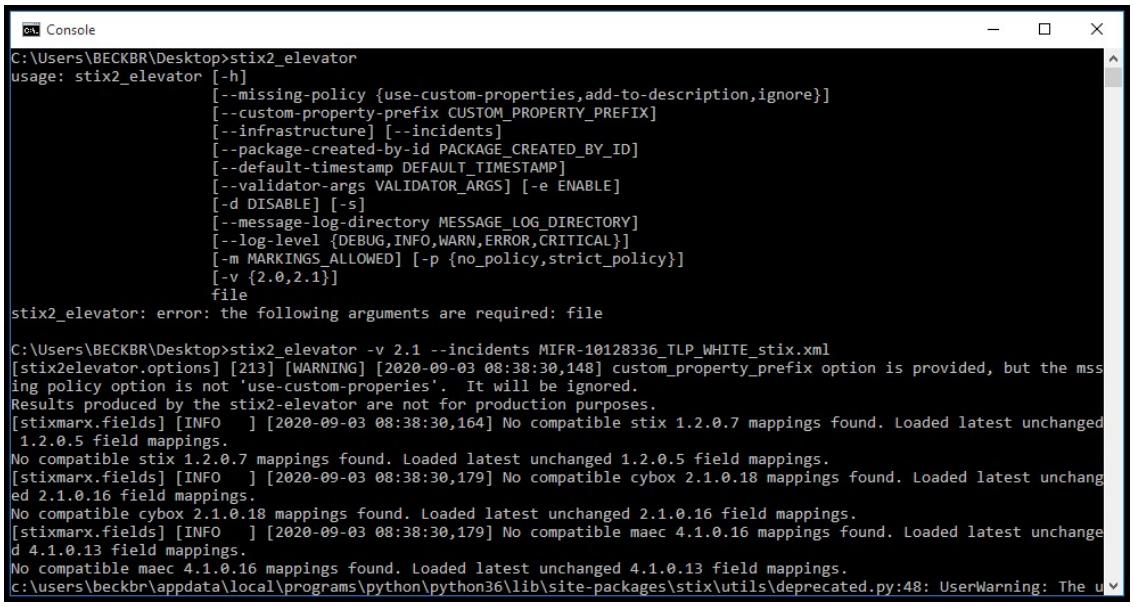


Figure 2. Sample output of the STIX2 Elevator.

Items that were lost during automated elevation include incidents and TTP objects, which are native in STIX1 but not directly supported in STIX2, antivirus detection names, some associated hashes, and even some object descriptions. During the elevation process, a number of incomplete/broken STIX2 objects are also generated. These broken elements need to be manually cleaned to allow the enrichment process to continue. This cleanup process involves removing orphaned objects (i.e., those with no relationships), repairing broken objects (i.e., missing timestamp strings, etc.) and finally replacing the temporary placeholder used for STIX1 incident objects with a more permanent STIX2 Attack Pattern object. Figure 3 illustrates this manual cleanup process in a series of four graphs. In this STIX2 representation, red icons denote malware objects, yellow icons represent indicator objects, and bright blue indicate attack patterns, as seen in the inset legend in Figure 3.

To visualize the STIX2 files, the Structured Threat Intelligence Graph (STIG) [5] is utilized. STIG was developed at Idaho National Laboratory to view, edit, and create STIX2 graphs in a visual drag and drop interface. The visual nature of STIG helps to reduce the human error that can occur when handling large STIX files—especially the cumbersome universally unique identifiers (UUIDs). The ability of STIG to easily change the graph layout allows analysts the ability to recognize patterns in the STIX data that may previously have gone unnoticed. Figure 4 shows an example of STIG displaying the STIX2 data from Unit42's Oilrig [6] analysis displayed using the breadth-first graph layout, which makes evident the criticality of the Attack Pattern object highlighted in red. If one can neutralize the highlighted Attack Pattern, the connected malware objects (red elements) would be defeated along with the majority of the Oilrig campaign.

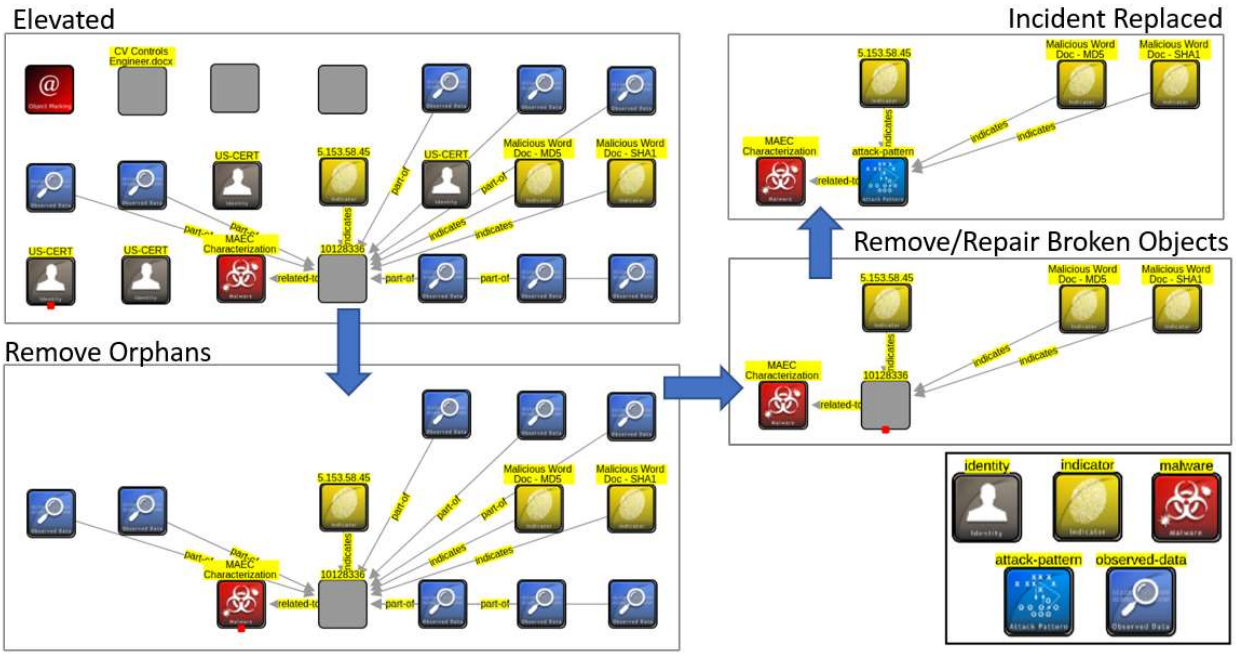


Figure 3. Manual cleanup process.

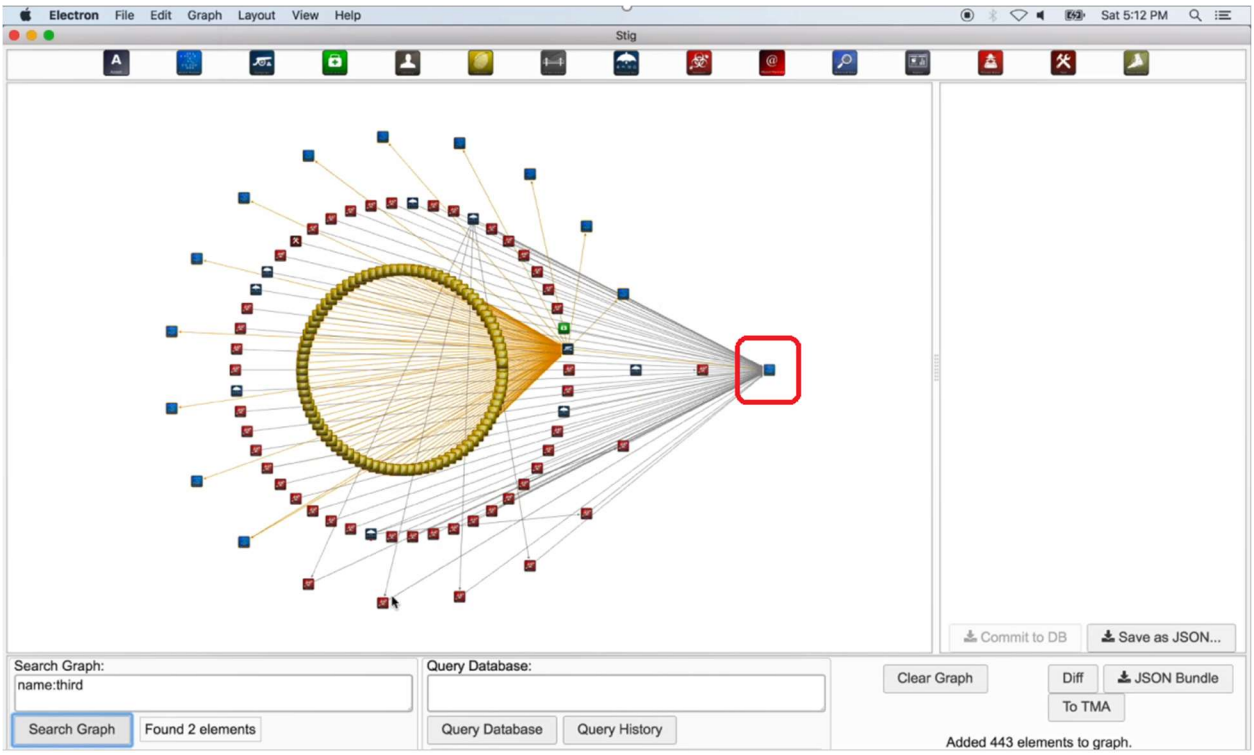


Figure 4. STIG displaying OilRig from Unit42.

With the manual cleanup process complete, further enrichment can continue. Parsing of the original STIX1 files and documents included in the original report allows for the creation of additional STIX objects. These manually created objects include malware objects enhanced with various antivirus identifiers, additional file hashes, and detailed object descriptions—all of which were lost as part of the

automated STIX2 elevation process. Figure 5 shows the JSON content of two of the manually created STIX2 objects, highlighting the antivirus names and additional hashes.

Armed now with a set of seven elevated and manually cleaned STIX2 bundles, as well as a large set of individual STIX2 objects that were manually created based on data existing in the original report, the process then involves the manual merging of both sets of STIX2 data, ensuring that the details and relationships from both data sets are captured in the single complete data set that results.


```
"type": "malware",
"id": "malware--1bc64917-fb43-4126-9711-032e2d2928b1",
"labels": [
    "adware"
],
"name": "MAEC Characterization of 722154a36f32ba10e98020a8ad7
"created_by_ref": "identity--5bb63797-09bb-4956-869f-52e09e6a
"created": "2017-06-13T13:17:27.575Z",
"modified": "2017-06-13T13:17:27.575Z",
"object_marking_refs": [],
"is_family": false,
"malware_types": [
    "McAfee - W97M/Downloader.cdg",
    "BitDefender - Trojan.GenericKD.12004346",
    "Microsoft Security Essentials - Trojan:O97M/Inoff.A",
    "Sophos - Troj/DocDl-JMD",
    "TrendMicro House Call - TROJ_RELSLODR.D",
    "TrendMicro - TROJ_RELSLODR.D",
    "Emsisoft - Trojan.GenericKD.12004346 (B)",
    "Ahnlab - DOC/Downloader",
    "ESET - DOC/TrojanDownloader.Agent.U trojan",
    "Ikarus - Trojan-Downloader.MSWord.Agent"
],
"spec_version": "2.1"
```

```
"type": "indicator",
"id": "indicator--d3e99adf-fc56-4ae7-92d3-67d0d6b1bbe2",
"labels": [
    "anomalous-activity"
],
"name": "CV Controls Engineer.docx - Hashes",
"pattern": "[file:hashes.'MD5' = '722154a36f32ba10e98020a8ad758a7a'] [file:hashes.'SHA-1' = '2872dcdf108563d16b6cf2ed383626861 [file:hashes.'SHA-256' = 'ac6c1df3895af63b864bb33bf30cb31 [file:hashes.'SHA-512' = 'e18d2e81dcbe4a5f334257d286e8880 [file:hashes.'SSDEEP' = '384:Dk5kSg2bPvHjd1coguI38aI2TUGI
"valid_from": "2017-06-13T17:29:51.309735Z",
"created_by_ref": "identity--2868c352-07d0-4797-90da-83c7508aed46",
"created": "2017-06-13T17:29:51.309Z",
"modified": "2017-06-13T17:29:51.309Z",
"object_marking_refs": [
    "marking-definition--613f2e26-407d-48c7-9eca-18e91df99dc9",
    "marking-definition--f75cba51-2ba5-4e0d-acd3-14d259b021d6"
],
"indicator_types": [
    "malware-artifacts"
],
"pattern_type": "stix",
"spec_version": "2.1"
```


Figure 5. Manually created malware object (left) and indicator object (right).

The final process in this enrichment procedure is manually grouping the associated elements from across the now complete seven STIX2 bundles. For this enrichment, the major groupings consisted of Spear-phishing, C2/Downloaders, Web Shells, SMB Redirect, Watering Hole, Obtain Persistence, Spear Phishing, PowerShell Scripts, and Tools. Additional objects were created and added to complete the final graph, including threat actor, campaign, vulnerability, and a few notional courses of action objects applicable to the identified vulnerability. The final and complete graph is seen in Figure 6.

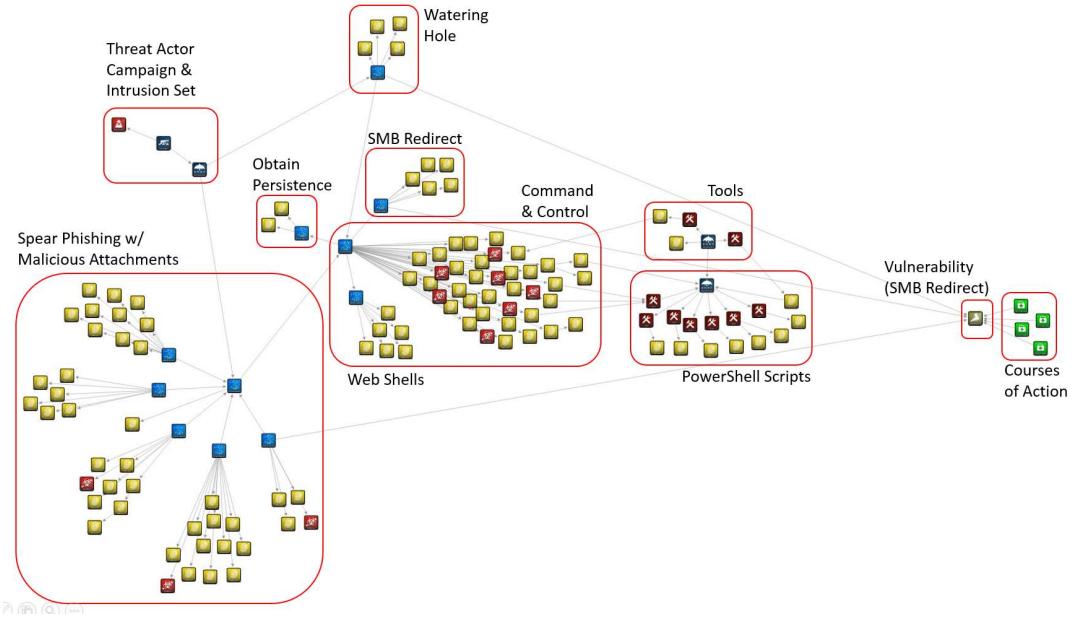


Figure 6. Final “Russia in the Grid” graph with groupings.

To quantify the quality of our enrichment process, we also developed a preliminary system to assign a numerical score to a STIX2 bundle. This scoring takes elements, including the number of relationships in a bundle, the overall number of objects present, and the length of descriptions provided for each object within a select STIX2 file, amongst others, into account. Scoring the STIX1 files immediately after being elevated to STIX2 provided an average score of 4.7 out of 13. Once the enrichment process was completed, the final "Russians in the Grid" STIX2 file scored a much higher score of 7.4 out of 13. This scoring system is still under development and will be used to judge the usefulness of a STIX2 bundle for inclusion as part of our machine learning predictive threat work.

While enriching the files from this report, a number of limitations were identified. One of the most difficult to resolve involved the STIX2 Elevator, which had difficulty processing custom fields, missing mappings, and missing vocabularies. Other elevator issues were found as well, which were related more to the content of the original STIX1 files. The original STIX1 files included fields and formatted URLs that did not align with the STIX2 specification. Most of the problems we encountered could have been avoided if the STIX files included in the report were written in STIX2 or STIX2.1 initially. The final process of grouping had some limitations as well. We chose to utilize the attack pattern object to group the related elements, as STIX2 does not have the concept of a generic grouping object. A grouping object has since been introduced in the STIX2.1 specification.

Fully transitioning from legacy XML-based STIX1 to a more modern JSON-based STIX2.1 would make a significant impact in the enrichment process we just described. This change would result in the ability to increase usability and shareability of cyber-threat intelligence. In the meantime, automating the manual processes detailed above would provide a significant benefit for future enrichment activities. Of the steps described, the manual cleanup process would lend itself well to being automated, as well as the manual merging as both are a "select, compare, repeat" process. The final grouping step and the manual creation process could be automated or at least accelerated through the use of machine learning and natural language processing techniques.

This enrichment research was conducted through the Geo Threat Observable (GTO) project, which was sponsored by the U.S. Department of Energy Office of Cybersecurity, Energy Security, and Emergency Response (DOE-CESER) Cybersecurity for Energy Delivery Systems (CEDS). Through this enrichment research, our team was able to take an already great report and increase its usefulness and effectiveness. This was accomplished by elevating the legacy STIX1 files to a modern STIX2 representation, as well as combining multiple related graphs into a single graph representation that can more easily be shared with other agencies, corporations, and asset owners—making it increasingly difficult for a threat actor to repeat an attack on other targets, which represents another attempt to try and stay one step ahead of adversaries.

Sources

[1] STIX Standard/Oasis: https://oasis-open.github.io/cti-documentation/stix/intro.html

[2] US-CERT Technical Alert (TA18-074A): https://us-cert.cisa.gov/ncas/alerts/TA18-074A.

[3] StixViz: https://github.com/STIXProject/stix-viz.

[4] STIX2-Elevator: https://stix2-elevator.readthedocs.io/en/latest/.

[5] STIG: https://github.com/idaholab/STIG.

[6] OilRig Analysis by Unit42:
https://github.com/pan-unit42/playbook_viewer/blob/master/playbook_json/oilrig.json.