Project

2021.10.05. Tuesday

* Subject : Analysis of cyber security incident IOC and resulting artifact extraction, data set construction and machine learning analysis

●Relevant Content Definition

- IoC(Indicator of Compromise) : After a cyber security incident occurs. it means related malicious codes which are collected, accident analysis reports, malicious code analysis information.
- Artifact : Information that can be used for cyber security incident analysis among the information included in the collected IoC
  (ex. malware name, hash value, DNS, IP, register information, file download/upload information, etc)
- IoC format : There are STIX, CYBOX, MISP, MAEC, etc. in the form of the provided IoC, and each IoC format is built in various languages such as JSON and XML.

1. Analysis of cyber security incident IOC and resulting artifact extraction, data set construction

1-1) Check the format of the provided data set

- Analyze the provided data set and check the format of IoC and the language to be used

1-2) Define extractable artifact information from a data set

- Define the artifact information that can be used for cyber security incident analysis among various information contained in IoC
- Information that can be obtained according to the format of IoC and language to be used can be found on the format of IoC website.

```
<stix:STIX_Header>
<stix:Observables cybox_major_version="2" cybox_minor_version="1" cybox_update_version="0">
    <cybox:Observable id="VMRayAnalyzer:Observable-b0aefe2b-dd8d-4aa0-b7b0-c35faab53f22">
        <cybox:Title>Process</cybox:Title>
        <cybox:Object id="VMRayAnalyzer:Process-0f6a2863-f018-47dd-abf5-4a03dadc36bb">
            <cybox:Properties xsi:type="ProcessObj:ProcessObjectType">
                <cyboxCommon:Custom_Properties>
                    <cyboxCommon:Property name="monitored_id">1</cyboxCommon:Property>
                </cyboxCommon:Custom_Properties>
                <ProcessObj:PID>2376</ProcessObj:PID>
                <ProcessObj:Name>whzqnu.exe</ProcessObj:Name>
                <ProcessObj:Parent_PID>1112</ProcessObj:Parent_PID>
                <ProcessObj:Image_Info>
                    <ProcessObj:File_Name>whzqnu.exe</ProcessObj:File_Name>
                    <ProcessObj:Command_Line>"C:\Users\5p5NrGJn0jS HALPmcxz\Desktop\whzqnu.exe" </ProcessObj:Command_Line>
                    <ProcessObj:Current_Directory>C:\Users\5p5NrGJn0jS HALPmcxz\Desktop\</ProcessObj:Current_Directory>
                    <ProcessObj:Path>c:\users\5p5nrgjn0js halpmcxz\desktop\whzqnu.exe</ProcessObj:Path>
                </ProcessObj:Image_Info>
            </cybox:Properties>
            <cybox:Related_Objects>
```

[Figure 1] An example of defining artifact information that can be used in the cyber security incident IoC. In IoC, process information is contained in stix:observavles-cybox:title, and PID, NAME, and Parent PID attribute values can be obtained from it.

- As shown in [Figure 1], find the artifact information included in IoC and define it as shown in <Table 1>

<Table 1> Cyber Security Incident IoC Artifact Information Definition

Attribute Name	Attribute Definition	Attribute Path
PID	PID information of processes used in cybersecurity incidents	'stix:Observables'→'cybox:Observable'→'cybox:object'→'cybox:properties'→'processobj:PIR'

1-3) Extraction of defined cyber security incident IoC artifact information and data set construction

- Extract the defined IoC attribute value

file_name	PID	Name	Parent PID
04ad737a6	2880	cscript.exe	1116
04ad737a6	1960	powershel	2880
9b86a50b3	2228	taskkill.exe	1884
9b86a50b3	2272	taskkill.exe	1884
9b86a50b3	2304	taskkill.exe	1884
hancitor-n	3024	convin~1.(	2940
hancitor-n	3032	explorer.ex	3024
hancitor-n	3040	cmd.exe	3032

[Figure 2] Example of extracting attribute values according to artifacts, In the process artifact of IoC, there are attribute names such as PID and Name, and the corresponding values can be extracted. (file_name is data file name)

- Since IoC contains various artifact information (process, register, malicious code information, file upload/download, network, etc.), it is necessary to extract attribute values from various artifacts in consideration of this situation.
- Based on the extracted attribute values, one data set is constructed using statistical and integrated analysis.

file_name	PID 수	대표 PID 이름	부모 PID 수	프로토콜명	포트번호	사용 파일 수
04ad737a6	2	cscript.exe	2	HTTP	42	5
9b86a50b3	3	taskkill.exe	1	HTTP	15	2
hancitor-n	3	convin~1.exe	3	FTP	10	1

[Figure 3] Example of data set,

By using the PID, Name, and parent PID obtained earlier, the number of PIDs and the number of parent PIDs were generated, and a single data set was constructed by obtaining the port number and the number of files used from other artifact information.