Situational Awareness Framework for Cyber Threat Intelligence Using LSTM based RNN

Mookyu Park, Jaehyeok Han, Moosung Park, and Kyungho Lee

Abstract— Since the Paris terrorist attacks in November 2015, new terrorist organizations such as ISIS are performing physical terrorism and cyber terrorism, unlike traditional terrorist attacks. These organizations use cyberspace as a vehicle to justify their attacks. To respond to these threats, the usability of cyber threat information is recovering. As a representative example, the United States developed the cyber threat information language STIX (Structured Threat Information eXpression) and its shared system TAXII (Trusted Automated Exchange of Indicator Information) centering on DHS (Department of Homeland Security) efforts are underway. However, it is difficult to extract cyber threat intelligence suitable for the situation and purpose with various information of cyberspace. This research proposes a cyber situational awareness framework to cope with future threats by recognizing environmental factors in cyberspace to overcome the limitations of current cyber threat information language. In addition, this paper classifies the components of FAIR (Factor Analysis of Information Risk) and STIX objects through RNN (Recurrent Neural Network) for the cyber situational awareness.

Research Keywords— Cyber Situational Awareness, Cyber Threat Intelligence, STIX, FAIR, RNN

1 INTRODUCTION

Recently, the cyber attack has been expanding in the cyberspace and has begun to have a negative impact on the real world. In particular, new terrorist groups such as ISIS have been carrying out cyberterrorism along with physical terrorism using bombs. Each country is conducting research and development on CTI (Cyber Threat Intelligence) in the aspect of national security to cope with these threats. A representative example is STIX(Structured Threat Information eXpression) which is the CTI standard language and TAX II (Trusted Automated Exchange of Indicator Information) which is the sharing system of DHS (Department of Homeland Security).

However, it is a reality that the decision maker relies on the existing heuristic technique or experience to make the information produced in the cyberspace to intelligence to suit the purpose of the decision maker. To overcome these limitations, this research proposes a situational awareness framework applying the risk measurement method FAIR (Factors Analysis of Information Risk). This paper matched STIX objects by learning each element of FAIR with RNN (Recurrent Neural Network) based on LSTM (Long Short-Term Memory) for cyber situational awareness.

2 RELATED WORKS & BACKGROUNDS

The initial cyber attacks were limited to cyberspace, such as hardware and software. However, recent cyber attacks such as ransomware, cryptocurrency exchange hacking, and information leaks are damaging to the real world by expanding in

- *Mookyu Park is with the Institute of Cyber Security & Privacy (ICSP), Korea University, Seongbuk-gu, Seoul, Republic of Korea, 02855. E-mail: ctupmk@korea.ac.kr.*
- *Jaehyeok Han is with the Institute of Cyber Security & Privacy (ICSP), Korea University, Seongbuk-gu, Seoul, Republic of Korea, 02855. E-mail: one01h@korea.ac.kr.*
- *Moosung Park is with the Agency for Defense Development(ADD), Songpa-gu, Seoul, Republic of Korea, 05661. E-mail: parkms@add.re.kr.*
- *Kyungho Lee (corresponding author) is with the Institute of Cyber Security & Privacy (ICSP), Korea University, Seongbuk-gu, Seoul, Republic of Korea, 02855. E-mail: kevinlee@korea.ac.kr.*

cyberspace. This section describes the cyber situational awareness and the FAIR model to respond to changing cyber attacks.

2.1 Cyber Situational Awareness (CSA)

Cyber situational awareness (CSA) means recognizing the environmental factors of time and space that occur in cyberspace and responding to future threats. The CSA includes all the factors that affect the cyber assets and hostile forces [1]. The typical framework of this CSA is based on Endsley's model.

Endsley's model consists of a process of understanding the elements of the environment(Level 1: Perception), understanding of meaning(Level 2: Comprehension) and state projection in the near future(Level 3: Projection) within the volume of time and space. The perception is a level that recognizes the status and attributes of related elements in the environment. The comprehension level is the step of synthesizing the elements of the perception level by analyzing and evaluating the situation. The projection level predicts how information analyzed at comprehension level will affect the state of the future operating environment over time. This is shown in Figure 1[2].

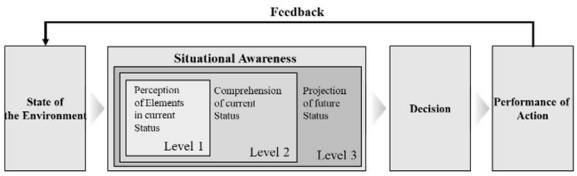

Fig. 1. Endsley's model consists of perception, comprehension, and projection, and makes decisions based on situational awareness.

2.2 FAIR (Factor Analysis of Information Risk)

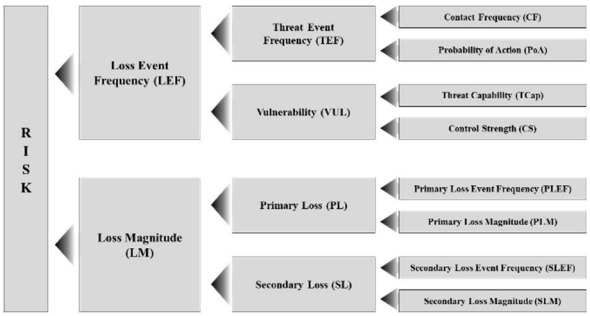

Fig. 2. The FAIR model is a suitable method for evaluating the intelligence according to the direction of the objective by measuring the risk as the detailed component of threats and assets.

The FAIR model measures risk with threats and assets and provides decision-making direction based on this. This model is divided into two components: Loss Event Frequency (LEF), which indicates the threat, and Loss Magnitude (LM), which indicates the loss of the asset. LEF is a variable indicating the frequency of the threat and is composed of Threat Event Frequency (TEF) and Vulnerabilities (VUL). LM consists of Primary Loss (PL) and Secondary (SL). This can be shown in Figure 2[3].

This paper proposes a method of utilizing cyber threat intelligence in cyber situational awareness by matching each element of FAIR and each objects of STIX.

3 CSA Systme Using STIX Objects

In CSA, decision making involves the probability of actual data, but it is highly influenced by human empirical aspects. In the case of intelligence extracted by purpose, human beings ultimately extract and combine this with situational awareness, there is a limit to objectivity. This section describes how to apply the deep learning method RNN as a way to overcome these limitations.

3.1 Concept

The domain structure of STIX, which is the standard language structure of CTI, consists of various elements such as Attack Pattern, Campaign, Course of Action, Identity, Indicator, Intrusion Set, Malware, Observed Data, Report, Threat Actor, Tool, and Vulnerability[4]. However, when converting the collected data to STIX, there is a limit to classify according to the definition of each element of STIX through empirical and cognitive aspects of the expert. This can act as an obstacle to the objectivity of intelligence. To overcome these limitations, this paper classified FAIR 's LEF elements into STIX objects by using RNN used in natural language processing. The reason for using only the LEF elements of the FAIR model in this research is that STIX focuses on threat rather than asset.

3.2 Matching Method Using RNN

The RNN is a method of sequentially processing information considering current input data and past input data. The RNN is a neural network that is intensively applied to natural language processing(NLP). Suppose that the value of the hidden layer at time t is h_t, the input x_t, coefficient matrix W at the same time t, the value h_{t-1} of the hidden layer at time $t-1$, and the matrix U indicating the

relationship between h_t and h_{t-1}, The transplantation is as follows.

$$h_t = \emptyset(Wx_t + Uh_{t-1})$$

The h input value x_t of the hidden layer is calculated at the output stage and W is updated based on the error[5]. This study uses RNN to determine the elements of the FAIR model to which the factor of CTI corresponds.

4 Result

This research matched each element of the FAIR model based on the domain object of STIX 2.0. In the case of learning data of RNN, the definition and description of Contact, Action, Threat Capability, and Control Strength corresponding to LEF of FAIR model were trained and each label was set. The input value uses the definition of each threat element of the STIX domain object. This paper has learned each element of the FAIR model using LSTM-based RNN that utilizes memory cells. The results for the learning data are shown in Figure 3.

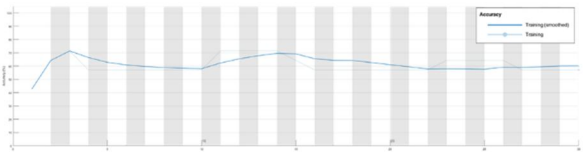

Fig. 3. The accuracy of the training data is 57.14%, but it can be solved if various definitions of FAIR's viewpoints can be entered.

Table 1. The STIX domain object can be classified into an action belonging to the TEF of the FAIR and a TCap belonging to the VUL.

STIX Domain Objects	Description	Element of FAIR
Attack Pattern	A type of Tactics, Techniques, and Procedures (TTP) that describes ways threat actors attempt to compromise targets	Action
Campaign	A grouping of adversarial behaviors that describes a set of malicious activities or attacks that occur over a period of time against a specific set of targets.	Threat Capability
Course of Action	An action taken to either prevent an attack or respond to an attack.	Threat Capability
Identity	Individuals, organizations, or groups, as well as classes of individuals, organizations, or groups.	Threat Capability
Indicator	Contains a pattern that can be used to detect suspicious or malicious cyber activity.	Action
Intrusion Set	A grouped set of adversarial behaviors and resources with common properties believed to be orchestrated by a single threat actor.	Action
Malware	A type of TTP, also known as malicious code and malicious software, used to compromise the confidentiality, integrity, or availability of a victim's data or system	Action
Observed Data	Conveys information observed on a system or network.	Action
Report	Collections of threat intelligence focused on one or more topics, such as a description of a threat actor, malware, or attack technique, including contextual details.	Threat Capability
Threat Actor	Individuals, groups, or organizations believed to be operating with malicious intent.	Action
Tool	Legitimate software that can be used by threat actors to perform attacks.	Action
Vulnerability	A mistake in software that can be directly used by a hacker to gain access to a system or network	Action

Using the LSTM-based RNN, the accuracy of the FAIR model is 57.14%. This is due to a lack of data on the definition and description of the FAIR model. In other words, although this study focuses on the basic definition of the FAIR model, this limitation can be resolved if there is a definition of the FAIR element for various scenarios. However, with this learning data, domain objects of STIX 2.0 can be classified. The results are shown in Table 1.

As a result of classifying STIX domain objects using LSTM based RNN, it was possible to classify them as elements belonging to TEF and VUL corresponding to LEF of FAIR. Attack Pattern, Indicator, Malware, and Threat Actor are classified as Action of TEF. Campaign, Course of Action, Identity and Report were classified as TCap belonging to VUL. Since the meaning of the action of the FAIR model is actually "a probability that a threat agent will act against an asset once contact occurs" and the meaning of the threat capability is "the probable level of a threat agent" 1 can be categorized as.

5 Conclusions

As the utilization of cyberspace increases, cyber attacks are being expanded to various targets such as personal information and financial information that may affect the real world as well as software and hardware. In addition, new terrorist organizations are pursuing terror through cyberspace. In response, many countries are investing in research and development of CTI that are effective in preventing cyber terrorism and cyber attacks in advance. However, in the case of CTI based on the intelligence concept, the empirical and cognitive aspects of the experts are strong, which can affect the objectivity of decision making. To overcome these limitations, this study classified the STIX objects through the deep learning method. This classification method can be used to effectively classify intelligence from the viewpoint of large capacity of CTI. It can also contribute to the decision-making and response of the national security system.

This study approaches the classification method of STIX which is a CTI standard in the cyber situational awareness system based on the FAIR model as a natural language processing viewpoint. However, since the FAIR model is a scenario-based risk measurement model, there is a limit to classify it by the existing definition alone. To solve these limitations, the future work will study the acquisition of training data through various scenarios and the situational awareness system of automation viewpoint through STIX

ACKNOWLEDGMENT

This work was supported by Defense Acquisition Program Administration and Agency for Defense Development under the contract. (UD060048AD)

REFERENCES

[1] R.A. Kemmerer, Cybaware: A Cyber Awareness Framework for Attack Analysis, Prediction, and Visualization.
[2] M.R.Endsley (1995). Toward a theory of situation awareness in dynamic systems. Human factors, 37(1), 32-64.
[3] J. A. Jones, An Introduction to Factor Analysis of Information Risk (Fair) 2005.
[4] S.Barnum(2012). Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX™). MITRE Corporation, 11, 1-22.
[5] A.Graves, A.R.Mohamed, & G.Hinton(2013, May). Speech recognition with deep recurrent neural networks. In *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on* (pp. 6645-6649). IEEE.

Mookyu Park received the B.S. degrees from Sejong University in 2014, Currently he is an Ph.D candidate in the Institute of Cyber Security & Privacy (ICSP) of Korea University since 2014. His research interests include Situational Awareness and Risk Management.

Jaehyeok Han received the B.S. degrees from University of Seoul in 2011, M.S. degrees in Information security from Korea University in 2016. Currently he is an Ph.D candidate in the Institute of Cyber Security & Privacy (ICSP) of Korea University. His research interests include Digital forensics, Data minig and Risk management.

Moosung Park received the B.S. degrees from Sogang University of in 1988, M.S. degrees from Sogang University in 1990, Ph.D degrees from Kwangwoon University in 2004. Currently he work for Agency for Defense Development. His research interests include Cyber Defense and Cyber Sitautional Awareness.

Kyungho Lee received the B.S. degrees from Sogang University of in 1989, M.S. degrees from Sogang University in 1997, Ph.D degrees from Korea University in 2009. Currently he is Associate Professor in the Department of Cyber Defense and the Institute of Cyber Security & Privacy (ICSP) of Korea University. His research interests include Risk Management, Information Security Consulting, Privacy Policy and Privacy Impact Assessment.