An Attribution of Cyberattack using Association Rule Mining (ARM)

Md Sahrom Abu[1] , Aswami Ariffin[4]
Malaysian Computer Emergency Response Team
Cybersecurity Malaysia, Cyberjaya
Selangor DE, Malaysia

Siti Rahayu Selamat[2], Robiah Yusof[3]
Faculty of Information Technology and Communication
Universiti Teknikal Malaysia Melaka
Durian Tunggal, Melaka, Malaysia

***Abstract*—With the rapid development of computer networks and information technology, an attacker has taken advantage to manipulate the situation to launch a complicated cyberattack. This complicated cyberattack causes a lot of problems among the organization because it requires an effective cyberattack attribution to mitigate and reduce the infection rate. Cyber Threat Intelligence (CTI) has gain wide coverage from the media due to its capability to provide CTI feeds from various data sources that can be used for cyberattack attribution. In this paper, we study the relationship of basic Indicator of Compromise (IOC) based on a network traffic dataset from a data mining approach. This dataset is obtained using a crawler that is deployed to pull security feed from Shadowserver. Then an association analysis method using Apriori Algorithm is implemented to extract rules that can discover interesting relationship between large sets of data items. Finally, the extracted rules are evaluated over the factor of interestingness measure of support, confidence and lift to quantify the value of association rules generated with Apriori Algorithm. By implementing the Apriori Algorithm in Shadowserver dataset, we discover some association rules among several IOC which can help attribute the cyberattack.**

Keywords*—CTI; association rule mining; Apriori Algorithm; attribution; interestingness measures

I. Introduction

With rapid development of computer networks and information technology such as internet connectivity, cloud storage and social media, various devices can easily connect to the internet. While this improvement has help internet users to access the latest information quickly, it also has bad consequences where an attacker can improve their tactic, technique and procedure (TTP) to launch a more complicated cyberattack. According to the statistic released by Malaysian Computer Emergency Response Team (MyCERT) as shown in Fig. 1, the number of malicious network activity, specifically on botnet in Malaysia had averagely surpassed 1 million unique IP infections per year [1].

This infection rate had caused a growing concern toward internet users in Malaysia because cybercriminals can manipulate the infected device for illegal activities. The infected machines can be used to deploy malware, initiate attacks on websites, steal personal information and mining cryptocurrencies. The number of infections rate is very alarming, and it causes a lot of problems among the organization because it requires an effective cyberattack attribution to mitigate and reduce the infection rate. Besides, this growing concern among internet users in Malaysia, Cyber Threat Intelligence (CTI) has gain wide coverage from the media due to its capability to provide CTI feeds from various data sources that can be used for cyberattack attribution. However, a proper process of voluminous data available in Cyber Threat Intelligence (CTI) is needed to achieve an effective cyberattack attribution.

Hence, the objective of this paper is to learn more about the relationship of basic Indicator of Compromise (IOC) using network traffic dataset from data mining approach. The network traffic dataset is obtain from Shadow server feed using a crawler. After that the extraction of rules to discover the interesting relationship between large sets of data items is conducted using an association analysis method. As a result, the implementation of association analysis method using Apriori Algorithm on Shadow server dataset can help to attribute the cyberattack based on useful information behind the association rules among several IOC.

The remaining of the paper is organized as follows: Section II presents the research background and related work based on association rules mining in CTI. Section III describes the proposed methodology that includes data collection using CTI feeds, data preprocessing and association analysis using the Apriori algorithm. While Section IV elaborates the rules extraction methods and represents the outcome of using interestingness measures to evaluate the rules generated. Finally, Section V provides a brief conclusion for this paper.

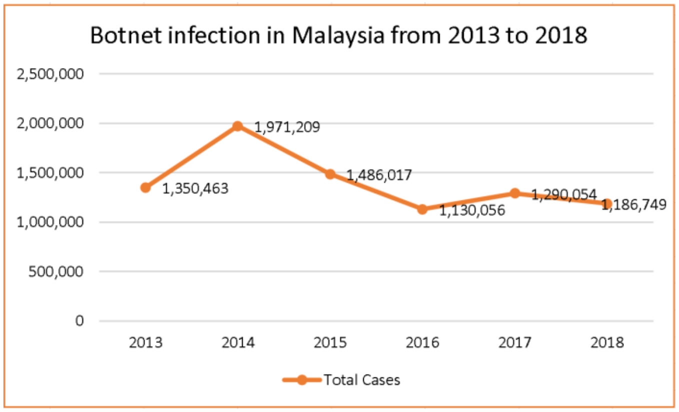

Fig. 1. Statistic of Botnet Infection in Malaysia.

II. Research Background and Related Works

A. Cyber Threat Intelligence (CTI) for Threat Attribution

There is no concrete definition to explain Cyber threat Intelligence (CTI) and it tends to change based on the working environment and business nature [2]. According to Gartner, CTI is evidence-based knowledge, including context, mechanisms, indicators, implications and actionable device, about an existing or emerging menace or hazard to asset that can be used to inform decisions regarding the subject's response to that menace or hazard [3]. While Pokorny et.al. [4] generally defines CTI as data that are collected from the operational environment, then it is processed and refined to produce information as shown in Fig. 2.

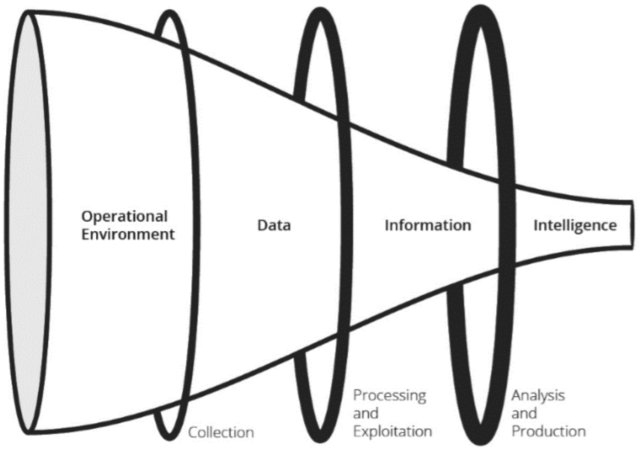

Fig. 2. Relationship of Data, Information and Intelligence [4].

This information is then analyzed and transformed into an actionable format that provides intelligence for threat attribution. Although defenses mechanism has evolved, attackers learned and upped their tactic, technique and procedure (TTP) by using fileless malware, code obfuscation, polymorphic designs and dynamic attack infrastructure that made basic IOC useless. So, Threat attribution is a demanding task, complicated and require a comprehensive intelligence or context [5].Threat attribution can be divided into four levels [6][7]. (1) Attribution to the specific hosts involved in the attack, (2) Attribution to the primary controlling host, (3) Attribution to the actual human actor, (4) Attribution to an organization with the specific intent to attack. These attribution levels are achievable through multiple techniques.

Wheeler [8] in his study, has described several techniques for cyberattack attribution that include tracing back based on log records, intrusion detection system, malware analysis and honeypots [9] as a guideline for a security analyst to identify the origin and threat actor behind the cyberattack. However, these traditional cyberattack attribution techniques have a limitation on discovering hidden knowledge beyond an IP address. The knowledge discovery beyond an IP address through an in-depth analysis of the problems from data sources especially focusing on association analysis is significant in helping security analysts to attribute the cyberattack effectively. Hence there has been a lot of research studies in the area of data mining to discover the useful and hidden knowledge among large groups of items or objects in transaction databases, relational databases, or other information repositories using Association Rule Mining (ARM) method.

B. Association Rule Mining (ARM)

Association Rule Mining (ARM) method has attracted many data mining researchers due to its capability to discover useful and interesting patterns from extensive, noisy, fuzzy and stochastic data. This method used to discover the relationship between variables in voluminous data. The strong relationship among variables is called association rules. These association rules contain two steps which are:

Frequent itemset identification (Support as the threshold): Find all frequent itemsets in a database that have transaction support above a predefined minimum threshold.

Rule generation (Confidence as the basic): These frequent itemsets use to generate the association rules that have confidence above a predefined minimum threshold.

Finding frequent itemsets in the database require more attention because of the difficulties involves searching for all the possible itemset combination. While rule generation in the second part is a straightforward task, these two parts can be represented as Equation 1.

$$P \rightarrow Q[s, c] \qquad (1)$$

Where:

The association between P and Q is whenever P appears Q also be likely to appear. P and Q may be a single condition or set of conditions. P is called the rule's antecedent part and Q is called a consequent part.

s; support is the probability that P and Q found together in a transaction.

c; confidence is the conditional probability that Q found in a transaction when P is present.

Currently, the most widely used algorithms in Association Rule Mining (ARM) is Apriori Algorithm. Agrawal [9] developed this algorithm to study customers' purchasing behavior in supermarkets where goods are often purchased together by customers. Besides, Apriori Algorithm also has been used in many areas of daily life successfully, including energy, recruitment, communication protocol, monitoring and network traffic behavior [10]. Hence the implementation of Apriori Algorithm in determining network traffic behavior can help security analysts to study attacker behavior in conducting cyberattack. Furthermore, the result regarding the attacker behavior from association rules generated using Apriori Algorithm can facilitate the security analyst on cyberattack attribution process.

1) Apriori algorithm: Apriori is an algorithm introduced by R. Agrawal and R Srikant in 1994 [9] and structured to focus on databases that consist of a large amount of transaction data. The basic Apriori algorithm utilised the bottom-up technique that makes it possible to extend the frequent subsets one item at a time which is termed as the candidate generation step. Further, clusters of candidates are evaluated against the data. The algorithm dismisses the process when no further relevant extensions are possible. The Association Rule Mining for Apriori Algorithm is defined as a two-step technique, namely, 1) generating the frequent itemset and 2) generating the rule. In the first step, it involves discovering all frequent item-sets that have support greater than or equal to a pre-determined minimum support count. While in the second step, it involves producing all the relevant Association Rules from frequent item-sets. In this step, it further involves on evaluating the Support and Confidence for all the rules and pruning the rules that do not reach the minimum support and minimum confidence threshold values.

This two-step technique can be elaborated using Table I.

TABLE. I. TRANSACTIONAL DATA

TID	TR-1	TR-2	TR-3	TR-4	TR-5
X	1	1	0	1	1
Y	1	1	1	0	1

Legend:
TID – Transaction ID
TR-Transactional
X-Itemset X
Y-Itemset Y

Table I containing the transactional data set that can be used to understand the basic terminology in Apriori Algorithm as the following:

1) Itemset: This is the collection of one or more items or products from the transactions. The term K-item-set denotes a set of k items or products. As illustrate in Table I, the K itemset represent by itemset $X = (x_1,x_2,...x_k)$ and itemset $Y=(y_1,y_2,..,y_k)$.

2) Support Count: The total number of occurrences of an itemset *X* and/or itemset *Y* is defined as support count

3) Support: Support [11] measure the usefulness of association rules. It is defined as a proportion of transactions in a dataset that contains the itemset. It measures the frequency of association. How many times *X* and *Y* involved in association rules occur together in the dataset. When the frequency of *X* and *Y* occurring at the same time is equal to or greater than the designated minimum support threshold, *X* and *Y* meet frequent itemsets. Support can be represented as Equation 2.

$$support\ (X \rightarrow Y) = \frac{Transactions\ containing\ both\ X\ and\ Y\ items}{Total\ number\ of\ transactions} \quad (2)$$

$$support\ (X \rightarrow Y) = \frac{3}{5} = 60\%$$

4) Confidence: Confidence is importance because it can indicate the strength or the reliability of an association rules [11]. It is defined as the ratio of the number of transactions that include all items in a frequent itemset to the number of transactions that include all items in the subset. It determines how frequently item *Y* occurs in the transaction that contains *X*. Confidence represented the conditional probability of an item as shown in Equation 3.

$$confidence\ (X \rightarrow Y) = \frac{Total\ number\ of\ transactions\ containing\ X\ and\ Y}{Total\ number\ of\ transactions\ containing\ item\ X} \quad (3)$$

$$confidence\ (X \rightarrow Y) = \frac{support\ (X, Y)}{support\ (X)}$$

$$confidence\ (X \rightarrow Y) = \frac{3}{4} = 75\%$$

5) Lift: The lift value is a measure of the importance of a rule [11]. The lift measures how frequent *X* and *Y* occur together than expected if they were statistically independent. Lift value 1 indicates *X* and *Y* are independent. The lift can be represented as Equation 4.

$$lift\ (X \rightarrow Y) = \frac{support\ (X \rightarrow Y)}{support\ (X) * support\ (Y)} \quad (4)$$

$$lift\ (X \rightarrow Y) = \frac{(\frac{3}{5})}{(\frac{4}{5}) * (\frac{4}{5})} = 0.384$$

The lift is a value between 0 and infinity:

a) If Lift (I) < 1, then X and Y are said to be interdependent on each other negatively.

b) If Lift (I) = 1, then X and Y did not find themselves dependent on each other and said they were independent.

c) If Lift (I) > 1, then X and Y appear together more often in the data and are said to depend on each other positively.

6) Frequent Itemset: The value of Support and Confidence determines the interestingness of the generated rule. This achieved by setting the minimum support and minimum confidence thresholds. The Item-sets whose support is greater than or equal to the specified minimum support threshold is defined as the frequent itemset.

Regardless of how the association rule is defined, it requires a suitable measure to achieve relevant and effective association rules by measuring the strongest dependencies between variables. For example, interestingness measures such as support, confidence, lift, correlation, and entropy have been used extensively to evaluate the interestingness of association rule.

2) Interestingness measure in ARM: Piatetsky-Shapiro introduced rule interestingness (RI) measures to evaluate the values of patterns [12] objectively. This measure can effectively quantify the correlation between the antecedents and the consequent for enormous association rules generated by ARM that can meet the aims of the researcher. The relevant and effective association rules are achievable through interestingness measure that includes objective measure and subjective measure [11]. The objective measures based on the statistical strengths or properties of the generated rules and subjective measures that are obtained from the user deduction or interest of their problem domain. Most of the research in the data mining field has use support and confidence as the de-facto "interestingness measures" for discovering relevant association rules [12]. However, support and confidence do not capture the correlation that exists between the antecedent and consequent of an association rule. There are several interestingness measures such as Laplace, Conviction, or Lift that can be used to fix this shortcoming. Among the three measures, Lift is the simplest yet the most powerful in capturing and representing the type of correlation exists between antecedent and consequent in association rule [13]. So, this paper will be used support and confidence as the de-factor interestingness measure and complement it with lift to evaluate the relevant association rule in facilitating cyber attack attribution.

III. Proposed Methodology

In this research, Association Rule Mining in CTI framework is performed using Apriori Algorithm as shown in Fig. 3. Fig. 3 illustrates the entire process of association rule mining in CTI framework that consists i) Preprocessing network traffic data, ii) Generating logical rules using Apriori algorithm and iii) Apply the generated rule to facilitate cyberattack attribution. The Apriori Algorithm can discover groups of items occurring frequently together in lots of transactions and such groups of items are called frequent itemsets. The association rule generated from this process is measured using support, confidence, and lift. Given a set of transaction, the problem of mining association rules is to generate all association rules that have support and confidence greater than the user-specified minimum support (called minsup) and minimum confidence (called minconf), respectively.

The implementation of Apriori Algorithm on Shadowserver dataset was done using R language. The capability of R language in performing statistical computing, data mining and graphics was optimize in this paper to process the filtered data and visualization.

A. Threat Intelligence Feeds

Data collection for this paper is limited to CTI feeds from OSINT that related to network intrusion activities. For this paper, OSINT CTI feeds from Shadowserver as shown in Fig. 4 has been chosen because it can provide various types of useful information and Indicators of Compromise (IoC) for cyberattack attribution [14]–[16]. The focus of this research is to gather CTI data that contain network resources from existing cyberattack.

Fig. 4 shows data collection process for Shadowserver dataset. Data collection started with collecting popular network resources such as the domain of search engines or government website, IP address of common DNS server and MD5 hash value of notorious malware. This network resource collected using crawler and APIs provide by Shadowserver. Then this data stored in excel storage before going through data preprocessing phase for data integration and data cleaning. The list of network resources and features from CTI data using APIs provide by Shadowserver is shown in Table II.

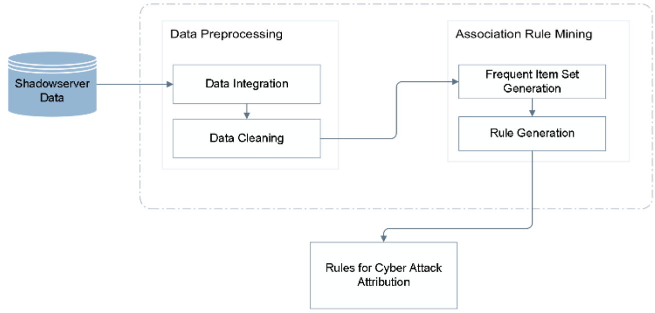

Fig. 3. Data Processing and Association Rule Analysis in CTI Framework.

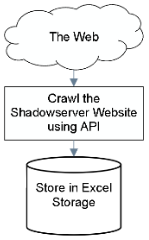

Fig. 4. Data Collection Process.

TABLE. II. The Detail Fields in Shadowserver feed

Data Source [Shadowserver]	Number of records	Features	Description
Dataset 1	334848	Timestamp	The start time of infection or attack occur
		Destination IP	malicious IPs used in the attack among the IPs reported in the threat report associated with a particular network resource.
		Source IP	
		Infection type	Malware type detect by antivirus

As far as the Shadowserver dataset is concerned, a record for network resources from this dataset is prepared with certain practical values correlated to precise network attributes [17]. Hence, in this research, the selected network resource for Shadowserver data include the following: the timestamp, the destination IP address, the source IP address, and the infection type detect by antivirus software. This dataset is obtained to discover significant knowledge behind raw data using ARM to facilitate cyberattack attribution process.

B. Data Preprocessing for CTI Feeds

Data scientists spent 80% of their effort in data preprocessing to produce intelligence from raw data that come in various formats and data types [18]. Data preprocessing can ensure the data in our possession are all fit, applicable and clean. It also plays a crucial role in increasing the accuracy of decision making by providing quality data. Data preprocessing phase consists of data integration and data cleaning that can be used to produce a clean and useful data before it can be used for ARM. This process is very important to make sure ARM can produce an effective association ruleset for cyberattack attribution. Data cleaning must be performed to identify potential issues with CTI feeds. With dirty, incomplete, noisy or otherwise "garbage in garbage out" data, the CTI framework unable to produce an effective cyberattack attribution. There are two steps are taken on performing a data preprocessing namely Data Integration and Data Cleaning. In Data Integration, all Shadowserver data merged into a single dataset; doing this results in duplicate IOCs such as IP addresses, hash value, domain name, URL and GeoIP. These duplicate IOCs are vital because it shows the correlation between different feeds. However, storing duplicate IOCs create redundant data. So, we need to perform data cleaning for this data set. While, in Data Cleaning, the cleaning is performed on the incomplete data set by filling in missing values or removing them altogether, along with eliminating noisy data and outliers. Other than that, identifying and repairing issues with the text that may cause data to become misaligned, such as embedded special characters, tabs or line breaks also perform. Once we get the clean data, we use these to do association rule mining to get the rules about network intrusion attack and analyze the meaning of the rules to facilitate cyberattack attribution.

C. *Association Rule Mining Algorithm in CTI Famework*

After the CTI feeds have been preprocessed for producing clean and useful data, the results will be used for association analysis to formulate an association ruleset. This association ruleset is to facilitate a cyber-attack attribution process in the CTI framework to produce an effective threat attribution. The association ruleset can assist security analysts in identifying the origin of the cyberattack and cyberattack attribution level.

As for the experimental setup using R, the configuration of threshold for minimum support value is 0.001 and the threshold of minimum confidence value is 0.5. There are 43 association rules meet this threshold configuration. Therefore to have a practical overview on the result generated, we use scatter plot as shown in Fig. 5 to visualize the association rules while Fig. 6 provides more information about these 43 rules using matrix visualization of grouped antecedents. The top-left corner plot of this matrix represents the most interesting rules based on lift measure.

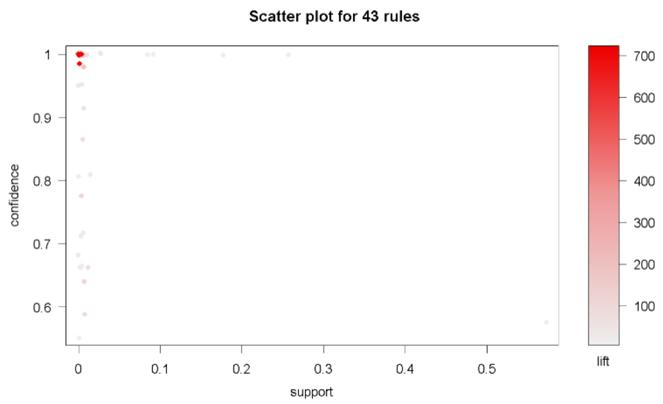

Fig. 5. The Illustration of 43 Rules in Scatter Plot with Minsup = 0.001 an Mincof = 0.5.

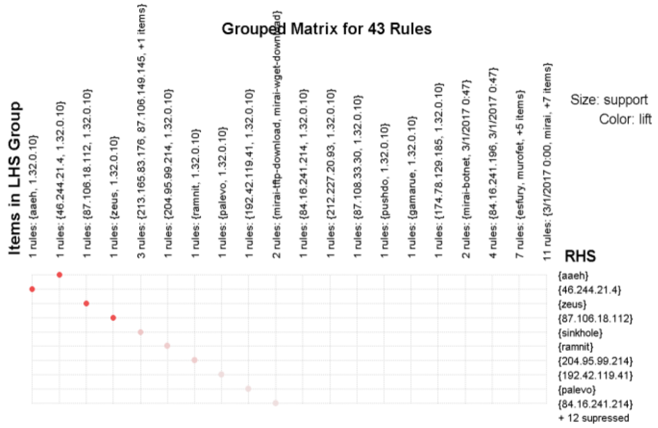

Fig. 6. The Visualization of 43 Rules in Group Matrix with Minsup = 0.001 and Mincof = 0.5.

Generally the experimental setup of this proposed methodology limit the value of support to 0.1 due to the large scale of dataset that we obtained and the confidence is configured between 0 to 1 which between the range allowed for confidence [19]. While the value of minimum support and the minimum confidence is adjusted manually to discover some specific and interesting rules from a large number of random rules [20]. Finally, the result from this experiment is presented in a table and sorted based on different measurement (support, confidence and lift) to analyze the relationship between IOC for this association rules.

IV. Association Rule Evaluation using Interestingness Measure

The number of association ruleset generated using ARM can be massive and even tricky for domain specialists to study and summarize the meaning behind the ruleset. Moreover, it is also impractical to sift through a broad set of rules containing noise and irrelevant rules. As a solution to this issue, interestingness measure can be used for filtering or ranking association rules. As discussed in Section 2, this paper will only focus on objective interestingness measure specifically using support, confidence and lift. While, the thresholds for minimum support (minsup) and minimum confidence (minconf) are manually defined by user [20][21][10].

A. *Evaluation by Lift*

Table III depicts the top five association rules with respect to lift measure. There are three categories to interpret the relationship of X / Y in lift measurement. If the lift is equal to 1, it means that X and Y are independent. If the lift is higher than 1, it means that X and Y are positively correlated. If the lift is lower than 1, it means that X and Y are negatively correlated. Based on Table III, we can see that the itemset 46.244.21.4, aaeh, 87.106.18.112 and zeus respectively have a positive correlation relationship with each other. While the item set 213.165.83.176 also has a positive correlation relationship with sinkhole. It shows that this IP is malicious and being sinkhole by an organization due to it malicious activity.

B. Evaluation by Support

Table IV shows top five association rules based on support with threshold configured as minsup = 0.15 and minconf = 0.15. Meanwhile, the visualization in Fig. 7 illustrates ten rules that satisfy this configuration.

Basically the top 10 rules in Fig. 7 sum up the combination of rules among Mirai, 212.61.180.100, 195.38.137.100 and Dorkbot which indicate there is a strong association among these four items that frequently occur together. Mirai is a malware that turns poorly conFig.d networked devices running Linux into botnets that can be used to launch a large-scale cyberattack that specifically targeted Internet of Thing (IOT) devices such as home routers, DVRs and webcams [22]. While, Dorkbot is a family of malware worms that spreads through instant messaging, USB drives, websites or social media channels like Facebook. Based on these two variants we can deduce that IP 212.61.180.100 belongs to an IOT device such as IP cameras or home router that has been infected through USB drives while IP 195.38.137.100 has been infected by dorkbot variant through malicious link that spread in social media such as Facebook and WhatsApp.

C. Evaluation by Confidence

Confidence measure can provide the most reliable association rule generated in experimental setup. Table V presented the top 5 most reliable rules with a threshold for minsup = 0.001 and minconf = 0.2 while Fig. 8 visualize the top 10 rules for this measurement.

The top 10 rules based on confidence measurement shows that high confidence rules usually related to sinkhole, zeus and aaeh.There is three IP that being sinkhole because being compromise and use by an attacker to launch malicious activity. IP 46.244.21.4 that associates with aaeh variant being used as a dropper to download other malicious code. While IP 87.106.18.112 that strongly associates with zeus being used to steal sensitive information that related to financial data.

TABLE. III. TOP 5 RULES BASED ON LIFT MEASURE WITH MINSUP = 0.001 AND MINCONF=0.5

Antecedent (*X*)	**Consequent (*Y*)**	**Support**	**Confidence**	**Lift**
46.244.21.4	aaeh	0	1	719.2
aaeh	46.244.21.4	0	1	719.2
87.106.18.112	zeus	0	1	702.2
zeus	87.106.18.112	0	0.98	702.2
213.165.83.176	sinkhole	0	1	231.7

TABLE. IV. TOP 5 RULES BASED ON SUPPORT MEASURE WITH MINSUP = 0.15 AND MINCONF=0.15

Antecedent (*X*)	**Consequent (*Y*)**	**Support**	**Confidence**	**Lift**
mirai	212.61.180.100	0.26	1.00	1.74
212.61.180.100	mirai	0.26	0.45	1.74
dorkbot	212.61.180.100	0.23	0.45	0.79
212.61.180.100	dorkbot	0.23	0.39	0.79
195.38.137.100	dorkbot	0.18	1.00	2.00

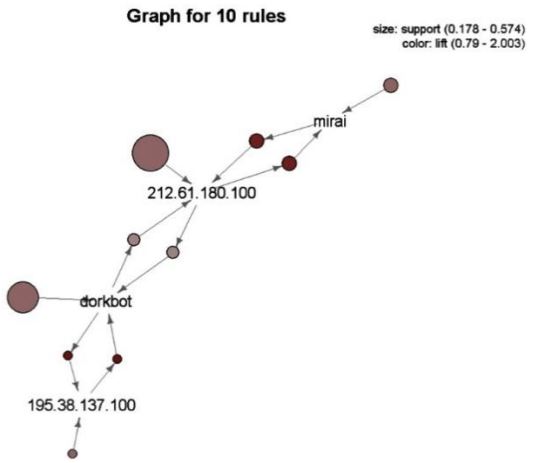

Fig. 7. Top 10 Rules based on Support Measurement with Minsup=0.15 and Minconf=0.15.

TABLE. V. TOP 5 RULES BASED ON CONFIDENCE MEASURE WITH MINSUP = 0.001 AND MINCONF=0.2

Antecedent (*X*)	**Consequent (*Y*)**	**Support**	**Confidence**	**Lift**
213.165.83.176	sinkhole	0	1	231.74
87.106.149.145	sinkhole	0	1	231.74
46.244.21.4	aaeh	0	1	719.19
aaeh	46.244.21.4	0	1	719.19
87.106.18.112	zeus	0	1	702.24

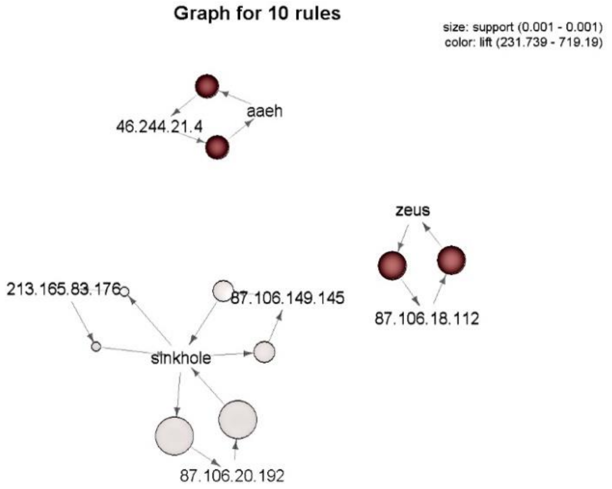

Fig. 8. Top 10 Rules based on Confidence Measurement with Minsup=0.001 and Minconf=0.2.

In this research, the same process is done using three different datasets from Shadowserver in order to prove the proposed work as shown in Table VI.

TABLE. VI. THE DETAIL FIELDS IN LEBAHNET FEED

Data Source [Shadowserver]	**Number of records**
Dataset 2	332874
Dataset 3	325730

Based on the results obtained, it discovers the strongest association rule indicates several malicious IP being targeted by cybercriminal is used to steal sensitive information that related to financial data. This association rule can help security analyst to focus on attack campaign and threat actor that related to financial attack. Apart from that, there is also an association rule that involved IOT devices such as IP camera and home router. This device most probably been compromised as a botnet to launch DDoS activity. From there, security analyst can focus on this IP for attack campaign and threat actor that actively involved in DDoS attacks. Security analyst also can further attribute this IP through pivoting and enrichment using third-party tools such as Passive DNS, Domain Tools IRIS and Maltego.

V. Conclusion and Future Works

Cyber threat intelligence provides a massive amount of raw data that contained useful information behind it. Association rule mining can help to discover significant knowledge behind this raw data to facilitate cyberattack attribution process in CTI framework.

In this paper, we employ Apriori algorithm to process CTI feed from Shadowserver dataset. Firstly, we explain the structure of Shadowserver dataset before going through data preprocessing process (data integration and data cleaning) using R language. Secondly, the Apriori Algorithm was explained in implementing the step of generating association rules. Finally, we evaluate the association rule using support and confidence as the de-factor in interestingness measure and complement it with lift to obtain the strongest association rules that reflect attacker characteristics when launching cyberattack. The finding of the experiment showed that the useful information about cyber-attack and attacker by association analysis can be discovered based on threat intelligence. In addition, the output data of association analysis can provide the information of cyber-attack relationship in cyber-attack attribution. For future work, more association rule algorithm and other statistical measures can be implemented to improve association ruleset effectiveness and accuracy in facilitating cyberattack attribution.

Acknowledgment

This study was kindly supported by The Ministry of Communications and Multimedia (KKMM), Cybersecurity Malaysia and Universiti Teknikal Malaysia Melaka (UTeM).

References

[1] MyCERT, “Malaysia Incident Statistic Report,” 2019. [Online]. Available: https://www.mycert.org.my/portal/statistics?id=b75e037d-6ee3-4d11-8169-66677d694932. [Accessed: 05-Jun-2019].

[2] Md Sahrom Abu and R. Y. , Siti Rahayu Selamat, Aswami Ariffin, “Cyber threat intelligence – Issue and challenges,” Indones. J. Electr. Eng. Comput. Sci., 2018.

[3] Gartner, “Definition: Threat Intelligence,” 2017. [Online]. Available: https://www.gartner.com/doc/2487216/definition-threat-intelligence. [Accessed: 10-Nov-2017].

[4] Z. Pokorny et al., The Threat Intelligence Handbook, Second Edition. 2019.

[5] L. Perry, B. Shapira, and R. Puzis, “NO-DOUBT: Attack attribution based on threat intelligence reports,” 2019 IEEE Int. Conf. Intell. Secur. Informatics, ISI 2019, pp. 80–85, 2019.

[6] J. Ryu and J. Na, “Security Requirement for Cyber Attack Traceback,” in 2008 Fourth International Conference on Networked Computing and Advanced Information Management, 2008, vol. 2, pp. 653–658.

[7] J. Hunker, B. Hutchinson, and J. Margulies, “Role and Challenges for Sufficient Cyber-Attack Attribution,” Inst. Inf. Infrastruct. Prot., pp. 5–10, 2008.

[8] D. A. Wheeler and G. N. Larsen, “Techniques for Cyber Attack Attribution,” Inst. Def. Anal. Rep., no. October, 2003.

[9] R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules (expanded version). Research Report IBM RJ 9839,” Proc. 20th Intl. Conf. VLDB, pp. 487--499, 1994.

[10] Y. Liu, K. Yu, X. Wu, Y. Shi, and Y. Tan, “Association rules mining analysis of app usage based on mobile traffic flow data,” 2018 IEEE 3rd Int. Conf. Big Data Anal. ICBDA 2018, pp. 55–60, 2018.

[11] C. Ju, F. Bao, C. Xu, and X. Fu, “A Novel Method of Interestingness Measures for Association Rules Mining Based on Profit,” vol. 2015, no. 2, 2015.

[12] M. Jalali-Heravi and O. R. Zaïane, “A study on interestingness measures for associative classifiers,” in Proceedings of the 2010 ACM Symposium on Applied Computing - SAC ’10, 2010, no. June, p. 1039.

[13] N. Hussein, A. Alashqur, and B. Sowan, “Using the interestingness measure lift to generate association rules,” J. Adv. Comput. Sci. Technol., vol. 4, no. 1, p. 156, 2015.

[14] X. Liao, K. Yuan, X. Wang, Z. Li, L. Xing, and R. Beyah, “Acing the IOC Game : Toward Automatic Discovery and Analysis of Open-Source Cyber Threat Intelligence,” pp. 755–766, 2016.

[15] Z. Zhu and T. Dumitras, “FeatureSmith: Automatically Engineering Features for Malware Detection by Mining the Security Literature,” Proc. 2016 ACM SIGSAC Conf. Comput. Commun. Secur. - CCS’16, pp. 767–778, 2016.

[16] C. Sabottke, O. Suciu, T. Dumitraș, C. Sabottke, and T. Dumitras, “Vulnerability Disclosure in the Age of Social Media: Exploiting Twitter for Predicting Real-World Exploits,” Proc. 24th USENIX Secur. Symp., 2015.

[17] S. Prakash and M. Vijayakumar, “An Effective Network Traffic Data Control Using Improved Apriori Rule Mining,” Circuits Syst., vol. 07, no. 10, pp. 3162–3173, 2016.

[18] J. Pérez, E. Iturbide, V. Olivares, M. Hidalgo, N. Almanza, and A. Martínez, “A data preparation methodology in data mining applied to mortality population databases,” Adv. Intell. Syst. Comput., vol. 353, pp. 1173–1182, 2015.

[19] A. Shah, “Association rule mining with modified apriori algorithm using top down approach,” Proc. 2016 2nd Int. Conf. Appl. Theor. Comput. Commun. Technol. iCATccT 2016, pp. 747–752, 2017.

[20] S. Mahmood, M. Shahbaz, and A. Guergachi, “Negative and positive association rules mining from text using frequent and infrequent itemsets,” Sci. World J., vol. 2014, 2014.

[21] L. Yan, Y. Ke, and W. Xiaofei, “Association Analysis Based on Mobile Traffic,” 2014 4th IEEE Int. Conf. Netw. Infrastruct. Digit. Content, 2014.

[22] P. Asghari, A. M. Rahmani, and H. H. S. Javadi, “Internet of Things applications: A systematic review,” Comput. Networks, vol. 148, no. 4, pp. 241–261, 2019.