J. Eng. Technol. Sci., Vol. 53, No. 6, 2021, 210610

Acquaintance Management Algorithm Based on the Multi-Class Risk-Cost Analysis for Collaborative Intrusion Detection Network

Yudha Purwanto[1,*], Kuspriyanto[2], Hendrawan[2] & Budi Rahardjo[2]

[1]School of Electrical Engineering, Telkom University,
Jalan Telekomunikasi, Bandung 40257, Indonesia
[2]School of Electrical Engineering and Informatics, Institut Teknologi Bandung,
Jalan Ganesha No. 10, Bandung 40132, Indonesia
*E-mail: omyudha@telkomuniversity.ac.id

Highlights:

- A collaborative intrusion detection network in a heterogeneous environment of IDS with classification capability was developed.
- Lower processing time of acquaintance management by the use of a merge risk-ordered acquaintance selection.
- High accuracy of intrusion decision based on the IDS acquaintance feedback by the use of multi-class risk-cost analysis in the acquaintance selection process.

Abstract. The collaborative intrusion detection network (CIDN) framework provides collaboration capability among intrusion detection systems (IDS). Collaboration selection is done by an acquaintance management algorithm. A recent study developed an effective acquaintance management algorithm by the use of binary risk analysis and greedy-selection-sort based methods. However, most algorithms do not pay attention to the possibility of wrong responses in multi-botnet attacks. The greedy-based acquaintance management algorithm also leads to a poor acquaintance selection processing time when there is a high number of IDS candidates. The growing number of advanced distributed denial of service (DDoS) attacks make acquaintance management potentially end up with an unreliable CIDN acquaintance list, resulting in low decision accuracy. This paper proposes an acquaintance management algorithm based on multi-class risk-cost analysis and merge-sort selection methods. The algorithm implements merge risk-ordered selection to reduce computation complexity. The simulation result showed the reliability of CIDN in reducing the acquaintance selection processing time decreased and increasing the decision accuracy.

Keywords: *acquaintance selection; collaborative; denial of service; intrusion detection; risk analysis.*

Received March 11th, 2020, 1st Revision December 8th, 2020, 2nd Revision February 28th, 2021, Accepted for publication April 29th, 2021.
 ISSN: 2337-5779,
DOI: 10.5614/j.eng.technol.sci.2021.53.6.10

1 Introduction

An intrusion detection system (IDS) is a technology that detects the existence of computer intrusions [1]. It is an essential part of the defensive system in network security. The detection system implements a signature based or anomaly based detection method. The capability of IDS in detection, prevention, and response capability has been investigated in Refs. [2,3]. However, with the growing number of advanced attacks, attacks are getting harder to detect. One of the reasons is that a single IDS has limited resources and knowledge to detect all attacks [4], especially large, coordinated botnet attacks [5].

A proposal to effectively resolve this problem is by implementing a collaborative intrusion detection network (CIDN), i.e., a collaboration framework among IDSs. CIDNs are widely implemented and have been studied in cloud computing [6], IoT [7,8], blockchain [9], and big data [10]. The purpose of these studies was to gain knowledge, information, and consultation among IDSs to improve IDS performance. Consultation based CIDN is a type of collaboration by sending consultation requests of observed data to be detected by the CIDN in order to improve the overall detection accuracy of the CIDN. The acquaintance management algorithm is a vital function of the framework. It selects the set of IDS that leads to the lowest overall risk-cost by evaluating the trustworthiness of each available IDS [11]. A greedy-selection-sort based algorithm was introduced in [4,12] to optimize the selection process instead of brute-force based selection. It uses a binary-based risk-cost case as the metric for trustworthiness in the selection process.

However, the greedy-selection-sort-based acquaintance management algorithm requires a relatively long acquaintance selection time when there is a large candidate list. This is because the computational complexity of the algorithm is $O(n^2)$ in a worst-case scenario and on average takes $T(n * l)$ running time. Moreover, our previous study [11] has shown that the use of a binary-based risk-cost analysis may produce an inaccurate risk-cost for multi-class detection cases, where the DDoS attack consists of more than one botnet or class of attack. This can lead to an unreliable set of selected collaborators, resulting in lower decision accuracy.

This research proposes an acquaintance management algorithm based on multi-class risk-cost analysis to reduce the acquaintance selection time and improve the accuracy of the risk-cost estimation. The proposed acquaintance management algorithm implements an ordered risk-cost approach to reduce its complexity. The trustworthiness of an IDS is evaluated by a multi-class risk-cost analysis to obtain an accurate risk-cost estimation. The effectiveness of the proposed acquaintance management algorithm was evaluated using the decision accuracy metric. The result showed that the proposed algorithm produces a more effective set of acquaintances with higher decision accuracy in less selection time compared to a comparative state-of-the-art algorithm.

This paper presents the following contributions. Firstly, this research developed a complete framework to simulate the process of a consultation-based CIDN that can simulate the flow of the detection process, collaboration updating, the selection process, and the feedback aggregation decision of the CIDN in order to improve its detection and classification accuracy. Secondly, this research proposes an acquaintance management algorithm that optimally selects a set of an acquaintances in less selection processing time with higher decision accuracy. Thirdly, this research developed a risk-cost analysis method based on multi IDS feedback by considering all possible consequences, including from wrong response decisions.

2 Related Work

A collaborative intrusion detection network (CIDN) is an overlay network that connects IDSs so that they can exchange information [4]. The collaboration works in three modes, i.e., information, knowledge, and consultation. In information mode, each IDS shares information about the detection result, such as alerts [13] and IP level security logs for a higher prediction ratio in proactive detection [14]. In knowledge mode, the new knowledge is shared among IDSs, such as new clusters [15] and new attack behaviors [16]. In consultation mode, the collaboration is done by sending consultation messages when the IDS has less confidence in the detection prediction, such as in [4]. The collaboration can be implemented in peer-to-peer [17], concentrated [18], and distributed [19] topologies.

In consultation-based CIDNs there are several important challenges in constructing an effective collaboration, such as collaboration management [20], incentive-based resource management [21, 22], malicious node detection [23,24], and consultation-request timing scenario [25]. The collaboration management algorithm selects a set of acquaintance IDS, where the trustworthiness can be estimated by several proposed evaluation parameters, such as satisfaction value [26], intrusion sensitivity [27], and risk-cost [4,11,28].

In risk-cost-based collaboration management, the IDS selects a set of acquaintances, resulting in the lowest overall risk-cost of the detection decision. Ref. [29] started consultation-based IDN research by trust-management to evaluate the behavior of IDN members. Its purpose is to select which IDS to collaborate with in order to improve the accuracy of attack detection. However, this study did not consider the possibility of rapid behavioral change as in malicious insider attacks. Ref. [28] proposed a consultation-based collaboration by aggregating IDS detection feedback. Risk-cost analysis based on IDS output was introduced to measure the trustworthiness of the IDS. In this case, an agent manager sends a consultation request to some IDS detection agents. This was studied further in Ref. [12], through collaboration management that not only selects but also manages the relationship in the CIDN. A greedy-selection-sort-based acquaintance management algorithm and binary risk-cost analysis were proposed, resulting in a relatively long acquaintance selection time when there is a large candidate list. This method potentially produces low decision accuracy when used in multi-botnet attacks, as the batch of consultation messages may consist of multi-class attacks. A recent study [11], proposed a multi-class risk-cost analysis, which leads to higher decision accuracy for multi-class attacks.

3 System Architecture

3.1 CIDN framework

Suppose there is an environment consisting of i number of IDS, $IDN = \{IDS_1, IDS_2, IDS_3, \dots, IDS_i\}$, parameterized by its performance, $IDS = [FP, FN, FoTP]$. Viewed from a CIDN point of view there are: *IDS_caller* ($IDS_s \in IDN$), the IDS in search of acquaintance, and several *called_IDS* ($IDS_j \in IDN; s \neq j$), IDSs other than IDS_s, that can potentially become IDS_s acquaintances. The roles of IDS_s and IDS_j are interchangeable, depending on the updating period of each IDS. From the IDS_s point of view, the available IDS_j that can collaborate with IDS_s first enter the probation list of IDS_s, $P^s = \{p_1, p_2, p_3, \dots p_j\}$, $P^s \subseteq IDN$, with $P^s = \{IDN\}/IDS_s$.

Acquaintance management in IDS_s evaluates the trustworthiness of each IDS_j in P^s. The goal is to select a set of IDS_j that has the lowest risk-cost from the acquaintance list. When the updating period arrives, the trustworthiness of IDS_j is evaluated by sending random test data $X_{test} = \{x^1, x^2, x^3, \dots, x^t\}$, where x is traffic data, to P^s. Each IDS_j in P^s then observes X_{test} with its detection method. The detection result Z_j^s from each IDS_j is then sent back to IDS_s. Based on a set of feedbacks $U_{test}^s = \{Z_1^s, Z_2^s, Z_3^s, \dots, Z_j^s\}$ from P^s, the IDS_s evaluates the detection performance $(H, F, FoTP)$ of each IDS_j. When the performance satisfies the IDS_s performance threshold, the IDS_j is moved to candidate list $C^s = \{c_1, c_2, c_3, \dots, c_n\}$, $C^s \subseteq P^s$.

However, not all IDS listed in C^s will be included in the collaboration, because the collaboration is looking for the lowest overall risk-cost of collaboration. Thus, IDS_j will be considered an acquaintance and included in acquaintance list $A^s =$

$\{a_1, a_2, a_3, \dots, a_i\}; A^s \subseteq C^s$ when IDS_j contributes to achieving the lowest overall risk-cost of acquaintance list $R(A)$. An illustration of this process is shown in Figure 1.

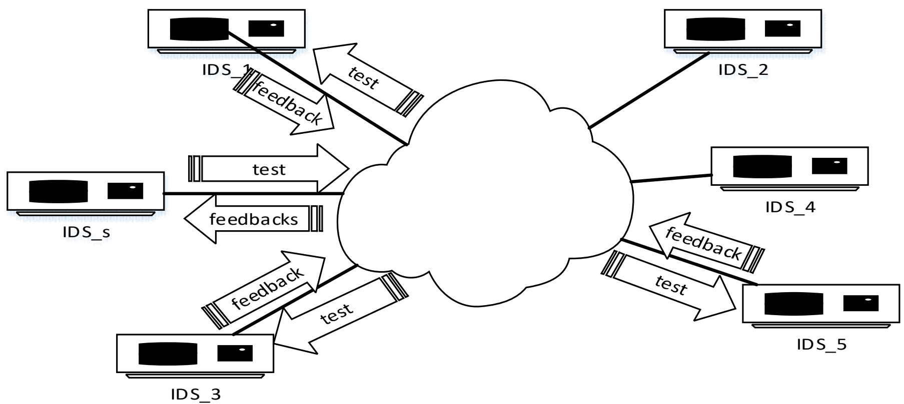


Figure 1 CIDN acquaintance selection process.

The collaboration in consultation-based CIDN is manifested in the consultation request of observed suspect data $X_{suspect}$, i.e the. message $CR(X_{suspect})$ from IDS_s to IDS_j in acquaintance list A^s. The messages communicated among CIDNs are shown in Table 1. The sending of consultation requests is not carried out on every observed data but based on the uncertainty of the IDS_s classification algorithm output. The classification output of any IDS can be analyzed from a confusion matrix such as in Figure 2.

Table 1 Messages exchanged between IDS.

Messages	Definition
$\boldsymbol{Join_{request}}$	Request from new ID_j to IDS_s to collaborate with
$\boldsymbol{X_{test}}$	Test data sent from IDS_s to IDS_j
$\boldsymbol{U_{test}}$	Detection reply of X_{test} from acquaintances to IDS_s
$\boldsymbol{CRX_{suspect}})$	Observed data sent from IDS_s to its acquaintances
$\boldsymbol{CF(U_{suspect})}$	Detection feedback of $X_{suspect}$ sent from acquaintances to IDS_s

The collaboration framework carries out four major functions. The first is detection and classification, including feature extraction by the use of feature selection and linear expansion; the second is the acquaintance management; the third is trust and consultation management; and the final function is making the feedback aggregation decision. The proposed framework is depicted in Figure 3.

	Predicted		
Actual	Normal	Attack A	Attack B
Normal	TN	FP	FP
Attack A	FN	TP	$FoTP$
Attack B	FN	$FoTP$	TP

Figure 2 Example of the classification confusion matrix.

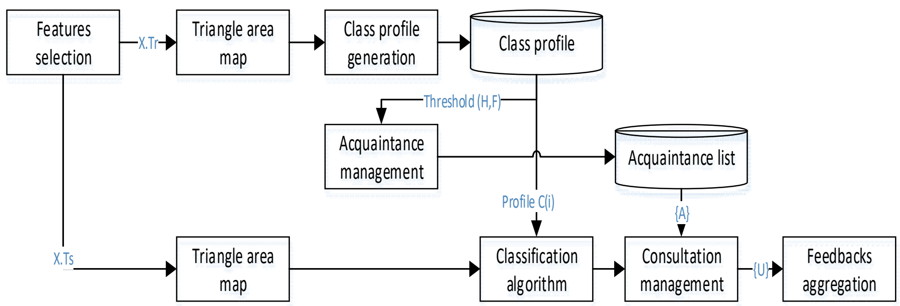


Figure 3 The framework of the proposed consultation-based CIDN model.

3.2 Acquaintance Management Algorithm

After receiving feedback items U_{test} from P, the IDS_s selects the candidate list by evaluating the feedback performance of Z_j from each IDS_j. From the available candidate list, acquaintance selection is done by evaluating the IDS_j in the candidate list that result in the lowest overall risk-cost. The overall risk-cost is estimated from the feedbacks risk-cost and the maintenance risk-cost.

This research proposes an acquaintance management algorithm to evaluate the overall risk-cost of the acquaintance list. The acquaintances are selected in a sequence of IDS_j, starting from the one with the lowest risk-cost $R(Z_j)$. The algorithm constructs the sequence by implementing a merge-sort method, which on average yields $O(n \log \log n)$ computation complexity. The present research developed the algorithm based on the research in [30], which was proved capable of providing high-quality candidates. The pseudo-code of the algorithm is depicted in Algorithm 1.

This research implements a feedback risk-cost analysis based on our previous research [11]. The analysis is carried out by considering the risk-cost of all consequence probabilities of the decision (δ) taken by IDS_s when given a set of U^s feedback items from CIDN. The risk-cost value can be seen as the estimated risk of cost or loss consequences when the system takes any decision according to observed traffic such as in Eq. (1). When the U^s input is a set of feedback, the risk value will depend on the marginal value of the parameters.

$$R(\delta) = (R(x = normal) + R(\delta | x = attack) \tag{1}$$

Algorithm 1. Acquaintance Management

at update event **do**
// send random X_{test} to P
// receives $U_{test} = \{Z_j\}$ from $IDS_j \epsilon P$
for all $IDS_j \epsilon P$ **do**
    if $t_j > t_{mature}$ **then**
        $P \leftarrow P \setminus IDS_j$
        if $H_j > H_s$ **and** $F_j > F_s$ **then**
            $C \leftarrow C \cup IDS_j$
        end if
    end if
end for
$A \leftarrow \{\ \}$
for all $IDS_i \epsilon C$ **do**
    $R_i \leftarrow R(Z_i)$
end for
$[index] \leftarrow sort(R_i)$
$C' \leftarrow$ Sort $IDS_i \epsilon C$ according to $index$
for $i = 1$ **to** $|C'|$ **do**
    $A \leftarrow A \cup IDS_i$
    if $R_{total}(A) < T$ **then**
        $T \leftarrow R_{total}(A)$
        $l = |A|$
        if $l = l_{max}$ **then**
            $j = |C'| + 1$
        end if
    else
    $A \leftarrow A \setminus IDS_i$
    end if
end for
return A

The parameter $p = P[X = 1]$ is the prior probability of an attack happening in the IDS. Decision δ can be in the form of a no_response when no attack is detected (normal traffic) or a response when an attack is detected. The risk analysis applies the product to several feedback items, $|A|$, from the IDS in the acquaintance list $\{A\}$. The risk-cost analysis follows the consequences from all possible decisions (δ), as shown in Figure 4.

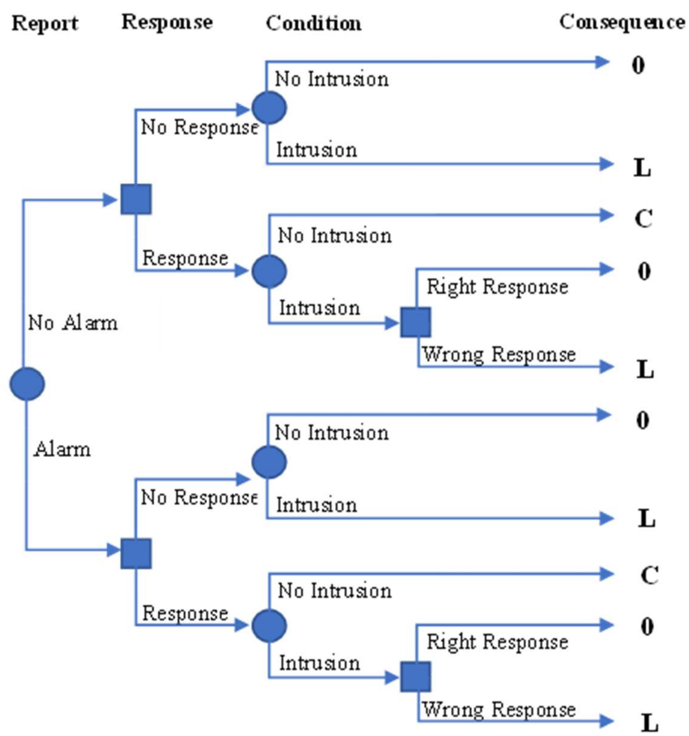


Figure 4 The decision tree of the expected risk-cost from all possible decisions (δ).

From each feedback Zi, the analysis gets information from a confusion matrix as shown in Figure 2. The loss consequence occurs in two possible response decisions, i.e. no_response and wrong_response, when intrusion occurs. A loss consequence of a no_response decision occurs when the detector output is a false negative (FN). A loss consequence of a wrong_response decision occurs when the attack is wrongly classified by the IDS so that the wrong_response is not able to stop the attack. This happens when the detector output is in false on true positive ($falseTP$) condition. A cost consequence occurs when the detector output is in false positive (FP) or true positive (TP) condition. However, this research only considered the cost consequence of false positive condition responses. The risk-cost formula for all consequences is listed in Table 2.

Table 2 Expected risk-cost consequences of the detection output.

Detector	Real condition	Expected risk-cost decision	
		No Response	**Response**
NoAlarm	No intrusion	0	$C * FN = C(((1-F)(1-p))$
	Intrusion	$L((1-H)p)$	$L((1-H)p * falseFN/FN)$
Alarm	No intrusion	0	$C * FP = C(F(1-p))$
	Intrusion	$L(H * p)$	$L(H * p * falseTP/TP)$

From Eq. (1) and Table 2, the following further analysis of risk-cost is obtained:

$$R(U) = \sum min\left(\left(L * p\prod_{i=1}^{|A|} H_i^{\alpha}(1-H_i)^{1-\alpha}\right), \left(\left(C * (1-p)\prod_{1}^{|A|} F_i^{\alpha}(1-F_i)^{1-\alpha}\right) + \left(L * p\prod_{i=1}^{|A|} H_i^{\alpha}(1-H_i)^{1-\alpha} \frac{false\, FN_i^{(1-\alpha)} false\, TP_i^{\alpha}}{FN_i^{(1-\alpha)}\, TP_i^{\alpha}}\right)\right)\right) \quad (2)$$

with $\alpha = \{0;\ \ if\ z = 0\ (normal)\ \ 1;\ \ if\ z \neq 0\ (attack)$

In CIDN, the overall risk-cost is influenced by two concerns. Firstly, the acquaintance risk-cost, which is the risk-cost of decisions from the obtained acquaintance list feedback, $R(U(A) = R(A)$. Secondly, the risk-cost due to the maintenance of collaboration. The value of the maintenance risk-cost can be seen from the number of resources allocated for the collaboration. The greater the number of collaborators, the more resources will be allocated for the computation and communication resources. For this reason, the risk-cost value of the maintenance process is formulated as a function of acquaintance list size ($|A|$). Thus, the total risk-cost is obtained by summing the CIDN feedback risk-cost and the maintenance risk-cost as follows:

$$R_{total} = R(A) + (|A| * \theta \quad (3)$$

4 Result and Analysis

This research applied supervised learning in the IDS detection method. The complete feature selection and normal profile generation processes are presented in [31] and [32]. The KDD Cup 99 dataset from [33] was applied to evaluate our model. For the learning phase, the DDoS dataset from KDD Cup 99 kddcup.data_10_percent was used as the basis for generating training data $X.Tr$. The DDoS dataset in kddcup_corrected was then used as the basis for testing dataset $X.Ts$, which was randomly generated in the testing phase. The simulation parameters are shown in Table 3.

Table 3 Simulation parameters.

Parameter	Value
$\boldsymbol{l_{min}}$	1
Threshold(H,F)	Based on IDS_s performance
$\boldsymbol{\frac{C}{L}}$	0.2
$\boldsymbol{W(X_{tes})}$	1000
$\boldsymbol{Mature}$	0.5
$\boldsymbol{t_{update}}$	50
$\boldsymbol{X_{test}}$ *composition*	Random
$\boldsymbol{\theta}$	0.001

A comparison between the proposed acquaintance management and a greedy-selection-sort algorithm based on [12] was conducted in the analysis phase. The comparison was done in terms of selection time, size of the acquaintance list $|A|$, and accuracy of the feedback aggregation decision of acquaintance feedback $(f_{mv}(U))$. The simulations were run within the scope of discrete event simulation with time parameter t symbolizing an activity [34].

The CIDN was modeled as IDS_caller (IDS_s) and called_IDS (IDS_j), where each IDS was parameterized by FP, FN and FoTP. For the analysis of the acquaintance management algorithm, 100 called_IDS were generated in the IDN set to represent the IDN environment. The analysis was done in four environments, i.e. $IDN_1 = [< 2\%, < 2\%, < 3\%]$, $IDN_2 = [< 2\%, < 5\%, 3 \leq FoTP < 6\%]$, $IDN_3 = [< 2\%, < 5\%, 6 \leq FoTP < 9\%]$, and $IDN_4 = [< 2\%, < 5\%, \geq 9\%]$. An IDS_caller was then generated for each environment, which was parameterized by $IDS_{caller} = [FP < 2, FN < 5, FoTP < 10]$. The analysis was done on the average of all IDS acquaintance management outputs.

4.1 Processing Time

The simulation result showed that the selection time of the proposed acquaintance management algorithm was lower than comparison [12]. This is in accordance with the algorithm's complexity of $O(n + n\ log\ n + n) = O(n\ log\ log\ n)$ for the proposed algorithm, compared to $O\left(n\frac{(n-1)}{2}\right) = O(n^2)$ for comparison. In the simulated scenario, the minimum number of acquaintances (l_{min}) increased along with the size of the candidate list ($l_{min} < n$). With the proposed algorithm, the minimum number of acquaintances (l_{min}) did not affect the selection processing time. This is because the selection iteration is executed only once. The comparison is shown in Figure 5.

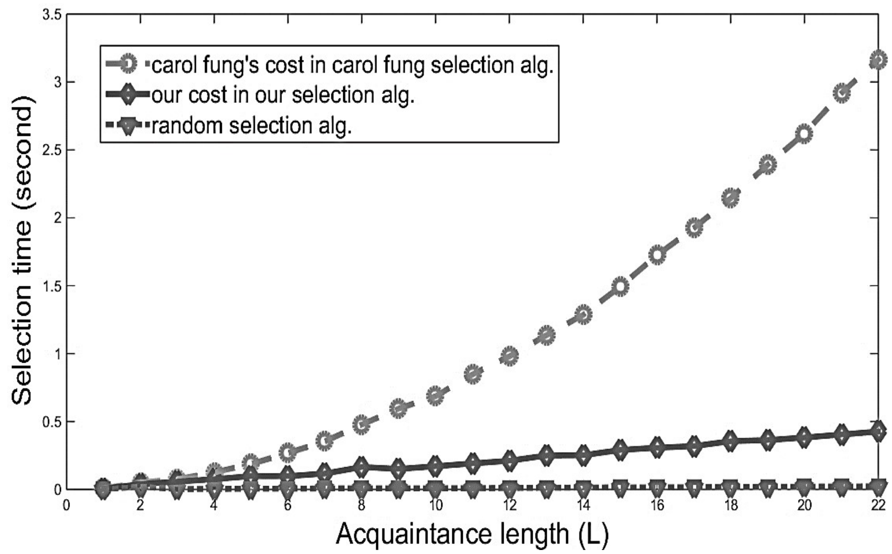
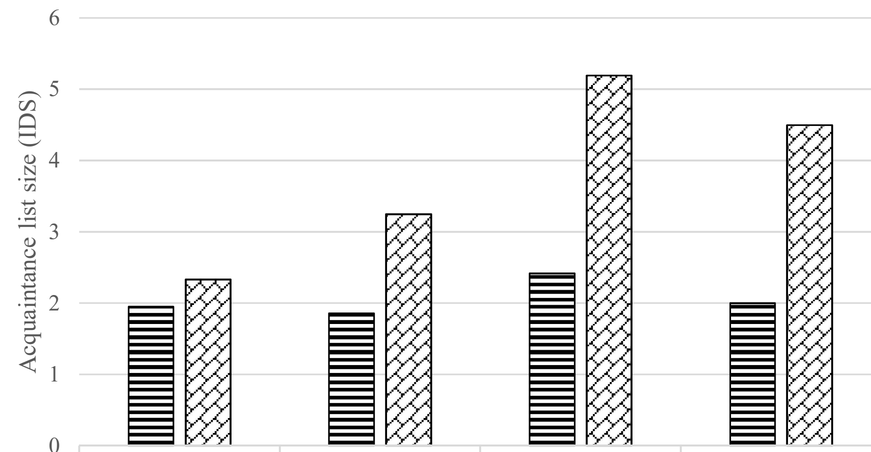


Figure 5 Acquaintance selection processing time comparison.

4.2 The Influence of Environmental Performance

In a CIDN environment with high classification performance such as IDN_1 there are more candidates and more combination options in acquaintance selection. Thus, the selection does not require too many members to achieve a low risk-cost value or high CIDN decision accuracy. However, in an environment with worse performance, such as IDN_2 or IDN_3, CIDN needs more $called\_IDS$ in the acquaintance list to gain high decision accuracy. The number of candidates in the candidate list also influences the obtained acquaintance list performance.

A smaller candidate list size leads to limited $called\_IDS$ options that can be collaborated with to improve CIDN performance. Thus, it decreases the acquaintance list size and the CIDN feedback decision accuracy, for example in IDN_4. The average acquaintance list size derived for every environment can be seen in Figure 6 where Alg. YP is the proposed algorithm and Alg. CF is comparison. The result is in line with the theoretical analysis discussed in the next section.

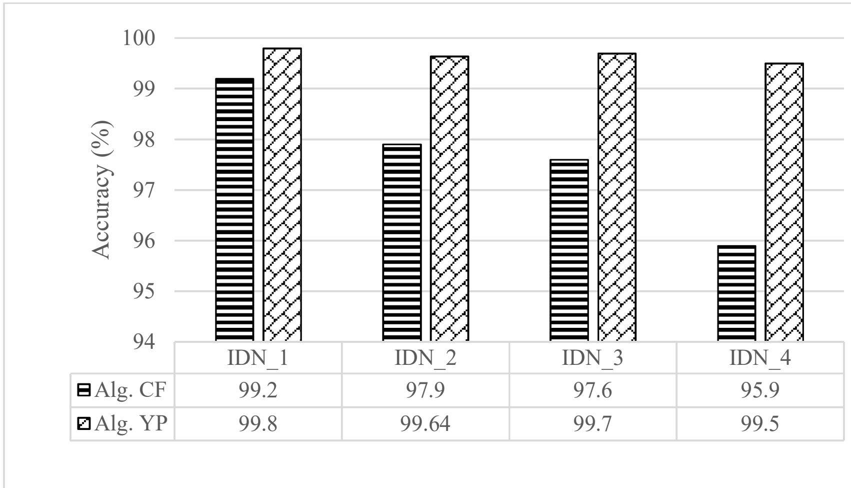


	IDN_1	IDN_2	IDN_3	IDN_4
Alg. CF	1.952380952	1.857142857	2.419354839	2
Alg. YP	2.333333333	3.25	5.193548387	4.5

Figure 6 Average acquaintance list size comparison.

From the accuracy performance result, the proposed acquaintance algorithm was able to provide an acquaintance list that produced better feedback decision accuracy than comparison. This is because the risk-cost analysis used in the proposed algorithm can accurately distinguish $called\_IDS$ trustworthiness. From the generated 100 IDS in the IDN set, the size of the probation and candidate list produced by both algorithms were the same in every scenario. However, as the risk-cost analysis from the proposed algorithm provided a more accurate estimation of risk-cost, the acquaintance list feedback decision accuracy was higher. This can be seen in the case example of feedback decision accuracy in every environment, as shown in Figure 7.

However, the proposed algorithm still had a drawback. The memory used in the selection process was higher because of the merge-sort method, which produces space complexity $O(n)$, i.e., higher than $O(1)$ for comparison. The higher number of acquaintance size also produces a higher number of consultation messages, which possibly burdens the network.



	IDN_1	IDN_2	IDN_3	IDN_4
Alg. CF	99.2	97.9	97.6	95.9
Alg. YP	99.8	99.64	99.7	99.5

Figure 7 Average acquaintance list accuracy comparison.

4.3 Theoretical Analysis

4.3.1 CIDN Decision Accuracy

Lemma:

> If $R(y)$ is a risk-cost value based on a stand-alone IDS decision y, and $R(A)$ is a risk-cost value based on feedback y_A from a group of $|A|$ IDS in CIDN, then $R(y) > R(A)$.

Proof:

From input data $X = \{x^1,\ x^2, x^3, \dots, x^t\}$, a stand-alone IDS detects and classifies X into $Y = \{y^1, y^2, y^3, \dots, y^t\}$. It has the probability of accurately classifying X in class C_1, which has $(\mu_{(y)}C1, \frac{1}{|A|}\sigma 2_{(y)}C1)$ as statistical profile, as formulated in Eq. (4):

$$P = \int \left(\frac{1}{\sqrt{2\pi}\sigma C1}\right) exp\ exp \left[-\frac{1}{2}\left(\frac{y-\mu C1}{\sigma C1}\right)^2\right] d_y \qquad (4)$$

In a CIDN with |A| number of acquaintances, each data x^t from $X_{test} = \{x^1, x^2, x^3, \dots, x^t\}$ will be sent to the acquaintances. Thus, the IDS will receive a set of feedback items $U = \{Z^1, Z^2, Z^3, \dots, Z^t\}$, with $Z^t = \{z_1^t, z_2^t, z_3^t, \dots, z_i^t\}$ and z_i^t is the feedback from $called\_IDS_i$ for data x^t. From the incoming data Z^t as the group of |A| data from the acquaintance list, the analysis is done on U by the use of the central limit theorem. From the statistical analysis on every data Z^t, i.e.,

$y_A = mean(Z^t)$, it constructs a C1 profile. Then, it has $(\mu_{(y_A)}C1, \frac{1}{|A|}\sigma^2{}_{(y_A)}C1)$ as C1 statistical profile. Thus, the probability of accurately classifying X in class C1 can be calculated as follows:

$$Q = \int \left(\frac{1}{\sqrt{2\pi}\,\sigma_{(y_A)}C1/\sqrt{\sqrt{|A|}}}\right) exp\left[-\frac{1}{2}\left(\frac{y_A - \mu_{(y_A)}C1}{\frac{\sigma_{(y_A)}C1}{\sqrt{\sqrt{|A|}}}}\right)^2\right] d_{y_A} \quad (5)$$

From P and Q analysis, the probability of accurately classifying X into C1 from CIDN decision (Q) is higher than from a stand-alone IDS (P) value. The Q value is classification accuracy, which is defined as $Acc = \frac{(TP+TN)}{|X|}$. A higher Q value means that the hit rate in CIDN is higher (H is directly proportional to TP) and the false rate is lower (F is inversely proportional to TN). Thus, if $Q > P$, then $R(A) < R(y)$, as shown by Eq. (2).

4.3.2 Acquaintance Management

Lemma:

If $R(z_s)$ is the risk-cost value based on the *IDS_caller* decision output, $R(z_1)$ is the risk-cost value based on feedback z^1 from $called\_IDS_1$, and $R(z_2)$ is the risk-cost value based on feedback z^2 from $called\_IDS_2$, where $R(z_1) < R(z_2)$, then the risk-cost from CIDN acquaintance $\{A_1\} = \{z_s, z_1)$ is lower than from acquaintance $\{A_2\} = \{z_s, z_2\}$; $R(A_1) < R(A_2)$.

Proof:

Suppose, a CIDN with |A| number of acquaintances applies a feedback aggregation decision $\delta = I(f_1(x) + f_2(x) + f_3(x) + \ldots + f_{|A|}(x))$, with $f_i(x)$ is the classification function of $called\_IDS_i$ in CIDN. Each f_i will have a classification error of $e(f_i(x))$, which aligns with risk-cost value (z_i) according to Eq. (2). Then, decision function δ will produce decision error $e = I(\delta \neq y)$ if $E = I(e(f_1(x)) + e(f_2(x)) + e(f_3(x)) + \ldots + e(f_i(x)) > T_e)$, where T_e is the error threshold.

For this reason, heuristic analysis selects $called\_IDS_i$, which has a lower $e(f_i(x))$ and produces a lower value of E. Suppose there are two $called\_IDS$ in candidate list $C = \{IDS_1, IDS_2\}$. From Bayes theorem for f_1 and f_2, if $P_e(f_1) =$

$P(e|f_1)$ is the probability of error in f_1 and $P_e(f_2) = P(e|f_2)$ is the one in f_2, then the probability of error Eqs. (6-8) are as follows:

$$P(f_1) = P(e \cap f_1)/P(f_1) \tag{6}$$

$$P(f_2) = P(e \cap f_2)/P(f_2) \tag{7}$$

$$P(e) = P(e \cap f_1) + P(e \cap f_2) = P(f_1)P(f_1) + P(f_2)P(f_2) \tag{8}$$

The probabilities of any error occurring in f_1 and f_2 are:

$$P(f_1|e) = P(e|f_1)P(f_1)/P(e) \tag{9}$$

$$P(f_2|e) = P(e|f_2)P(f_2)/P(e) \tag{10}$$

Heuristically, it is clear that if $P(f_1) > P(f_2)$, then $P(e) > P(e)$. By using a sorting method in acquaintance selection, a lower $e(f_i(x))$ value will result in a lower probability of aggregation decision error $P_E(\delta)$. Thus, it will have a lower risk-cost $R(A)$. In the case of $R(z_1) < R(z_2)$, the result has $R(A_1) < R(A_2)$.

5 Conclusion and Future Work

The proposed acquaintance management algorithm utilizes a sequence of sorted risk-cost candidates in the acquaintance selection process. Compared to a state-of-the-art algorithm, the proposed algorithm provides a reduced selection processing time and higher CIDN decision accuracy. In the proposed algorithm, the overall risk-cost value is estimated only in one iteration during the acquaintance selection process. This reduces the algorithm's computation complexity to $O\ (n\ log\ n)$, i.e., lower than $O\ (n^2)$ for the comparative algorithm. By using our previous risk-cost analysis in the proposed algorithm, it was proven to be able to select the acquaintance list that leads to the lowest overall risk-cost value and to a 2.7 percent higher CIDN decision accuracy on average. However, as a consequence of the implementation of merge-sort, the space complexity of the algorithm is higher ($O(n)$) compared to that of the comparative algorithm ($O(1)$).

For a better understanding of CIDN, the necessity of autonomous decision-making in CIDN will be investigated further in a future study by the use of a cooperative multi-agent model. Also, the necessity of resource management research, which directly concerns consultation management and incentives for collaboration, will be part of our future research.

References

[1] Mirkovic, J. & Reiher, P., *A Taxonomy of DDoS Attack and DDoS Defense Mechanism*, ACM SIGCOMM Computer and Communication Review, **34**(2), pp. 39-53, April 2004.

[2] Bhuyan, M.H., Bhattacharyya, D.K. & Kalita, J.K., *Network Anomaly Detection: Methods, Systems and Tools*, IEEE Communications Surveys & Tutorials, **16**(1), pp. 303-336, February 2014.

[3] Purwanto, Y., Kuspriyanto, Hendrawan & Rahardjo, B., *Traffic Anomaly Detection in DDoS Flooding Attack*, in International Conference on Telecommunication Systems, Services, and Applications, Bali, Indonesia, October 2014.

[4] Fung, C. & Zhu, Q., *FACID: A Trust-based Collaborative Decision Framework for Intrusion Detection Networks*, Elsevier Ad Hoc Networks Journal (ADHOC), **53**, pp. 17-31, December 2016.

[5] Soldo, F., *Predicting Future Attacks Data Analysis of Dshield Data Set*, Technical Report, 2009. http://www.ece.uci.edu/~athina/PAPERS/ dshield-analysis-tr. pdf. (December 2019)

[6] Le, D.N., Bhatt, C. & Madhukar, M., *Security Designs for the Cloud, IoT, and Social Networking*, John Wiley & Sons, Inc and Scrivener Publishing LLC, 2019.

[7] Benkhelifa, E., Welsh, T. & Hamouda, W., *A Critical Review of Practices and Challenges in Intrusion Detection Systems for IoT: Toward Universal and Resilient Systems*, IEEE Communications Surveys & Tutorials, **20**(4), pp. 3496-3509, 2018.

[8] Nguyen, T.G., Phan, T.V., Nguyen, B.T., Baig, Z.A. & Sanguanpong, S., *SeArch: A Collaborative and Intelligent NIDS Architecture for SDN-Based Cloud IoT Networks*, IEEE Access, **7**, pp. 107678-107694, 2019.

[9] Meng, W., Tischhauser, E.W., Wang, Q., Wang, Y. & Han, J., *When Intrusion Detection Meets Blockchain Technology: A Review*, IEEE Access, **6**, pp. 10179-10188, 2018.

[10] Tan, Z., Nagar, U.T., He, X., Nanda, P., Liu, R.P., Wang, S. & Hu, J., *Enhancing Big Data Security with Collaborative Intrusion Detection*, IEEE Cloud Computing, **1**(3), pp. 27-33, 2014.

[11] Purwanto, Y., Kuspriyanto, Hendrawan & Rahardjo, B., *Cost Analysis for Classification-based Autonomous Response System*, International Journal of Network Security, **20**(1), pp. 121-130, January 2018.

[12] Fung, C., *Design and Management of Collaborative Intrusion Detection Network*, Doctor of Philosophy Thesis of University of Waterloo, Ontario, Canada, 2013.

[13] Vasilomanolakis, E. & Mühlhäuser, M., *Detection and Mitigation of Monitor Identification Attacks in Collaborative Intrusion Detection*

Systems, International Journal of Network Management, **29**(2), e2059, 2019. DOI: 10.1002/nem.2059.

[14] Rezapour, A. & Tzeng, W.G., *A Robust Algorithm for Predicting Attacks Using Collaborative Security Logs*, Journal of Information Science and Engineering, **36**(3), pp. 597-619, 2020.

[15] Abdurrazaq, M.N.K., Trilaksono, B.R. & Rahardjo, B., *DIDS Using Cooperative Agents Based on Ant Colony Clustering*, Journal of ICT Research and Applications, **8**(3), pp. 213-233, 2015.

[16] Hung, J.C., *The Behavior-based Intrusion Detection and Response System for the Internet Worm*, Journal of Internet Technology, **4**(4), pp. 247-254, October 2003.

[17] Meng, W., Li, W., Yang, L.T. & Li, P., *Enhancing Challenge-Based Collaborative Intrusion Detection Networks Against Insider Attacks Using Blockchain*, International Journal of Information Security, **19**(3), pp. 279-290, 2019. DOI: 10.1007/s10207-019-00462-x.

[18] Zhou, C.V., Leckie, C., Karunasekera, S. & Peng, T., *A Self-healing, Self-protecting Collaborative Intrusion Detection Architecture to Traceback Fast-flux Phishing Domains*, in IEEE Workshop on Autonomic Communication and Network Management (ACNM 2008), Salvador da Bahia, Brazil, April 2008.

[19] Li, X., *Collaborative Intrusion Detection Method for Marine Distributed Network, Journal of Coastal Research*, Advances in Sustainable Port and Ocean Engineering, (Special Issue 83), pp. 57-61, 2018.

[20] Kanth, V., McAbee, A., Tummala, M. & McEachen, J.C., *Collaborative Intrusion Detection leveraging Blockchain and Pluggable*, in The 53rd Hawaii International Conference on System Sciences, Hawaii, 2020.

[21] Zhu, Q., Fung, C., Boutaba, R. & Basar, T., *GUIDEX: A Game-theoretic Incentive-based Mechanism for Intrusion Detection Networks*, IEEE Journal on Selected Areas in Communications, **30**(11), pp. 2220-2230, December 2009.

[22] Rezapour, A. & Tzeng, W.G., *A Robust Intrusion Detection Network Using Thresholdless Trust Management System with Incentive Design*, in Beyah R., Chang B., Li Y., Zhu S. (Eds.), *Security and Privacy in Communication Networks*, SecureComm 2018, in Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, **255**, Springer Verlag, 2018. DOI: 10.1007/978-3-030-01704-0_8.

[23] Li, W., Meng, W. & Kwok, L.F., *Investigating the Influence of Special On-Off Attacks on Challenge-Based Collaborative Intrusion Detection Networks*, Future Internet, **10**(1), pp. 1-16, 2018.

[24] Li, W. & Kwok, L.F., *Challenge-Based Collaborative Intrusion Detection Networks under Passive Message Fingerprint Attack: A Further Analysis*, Journal of Information Security and Applications, **47**, pp. 1-7, 2019.

[25] Purwanto, Y., Kuspriyanto, Hendrawan & Rahardjo, B., *Consultation Request Algorithm in Distance Based Intrusion Detection Network*, in International Conference on Satellite Technology, Bandung, Indonesia, October 2018.

[26] Fung, C. J., Zhang, J., Aib, I. & Boutaba, R., *Dirichlet-based Trust Management for Effective Collaborative Intrusion Detection Networks*, IEEE Transaction on Network and Service Management, **8**(2), pp. 79-91, June 2011.

[27] Li, W. & Meng, W., *Enhancing Collaborative Intrusion Detection Networks Using Intrusion Sensitivity in Detecting Pollution Attacks*, Information and Computer Security, **24**(3), pp. 265-276, 2016.

[28] Orfila, A., Carbo, J. & Ribagorda, A., *Autonomous Decision on Intrusion Detection with Trained BDI Agents*, Computer Communications Journal, **31**(9), pp. 1803-1813, June 2008.

[29] Duma, C., Karresand, M., Shahmehri, N. & Caronni, G., *A Trust-Aware, P2P-based Overlay for Intrusion Detection*, in International Conference on Database and Expert Systems Applications, Krakow, Poland, September 2006.

[30] Fung, C. J., Zhang, J. & Boutaba, R., *Effective Acquaintance Management Based on Bayesian Learning for Distributed Intrusion Detection Networks*, IEEE Transactions on Network and Service Management, **9**(3), pp. 320-332, September 2012.

[31] Purwanto, Y., Kuspriyanto, Hendrawan & Rahardjo, B., *Multistage Process to Decrease Processing Time in Intrusion Prevention System*, in International Conference on Wireless and Telematics, Palembang, Indonesia, July 2017.

[32] Purwanto, Y., Kuspriyanto, Hendrawan & Rahardjo, B., *Minimal Triangle Area Mahalanobis Distance for Stream Homogeneous Group-based DDoS Classification*, International Journal on Electrical Engineering and Informatic, **10**(2), pp. 369-383, June 2018.

[33] KDD Cup 99, Available on: http://kdd.ics.uci.edu/databases/kddcup, 1 October 1999. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html. (March 28, 2018)

[34] Ziegler, B.P., *Multifacetted Modelling and Discrete Event Simulation*, Orlando: Academic Press, London, 1984.