Structured Whitelist Generation in SCADA Network using PrefixSpan Algorithm

Woo-Suk Jung[1], Jeong-Han Yun[2], Sin-Kyu Kim[2], Kyu-Seok Shim[1] and Myung-Sup Kim[1]
Dept. of Computer and Information Science Korea University, Sejong, Korea[1]
National Security Research Institute, Daejeon, Korea[2]
{hary5832, kusuk007, tmskim}@korea.ac.kr[1], {dolgam, skkim}@nsr.re.kr[2]

***Abstract*— SCADA system works in repeated or periodic used of only limited communication devices. Because of this feature, whitelist based security techniques are widely used and access restriction method using whitelist based static ACL is most commonly applied in security field. Static ACL have advantages in security, but their expressiveness is too simple to express communication using dynamic allocated port. In addition, it does not reflect all the communication characteristics of the control device, and the generated static ACL should always be open regardless of the frequent use. We propose a structured ACL that extends the fixed generation sequence information between the communication and communication-specific periodicity to reflect the mechanical and repetitive communication characteristics of the SCADA system in the static ACL. We demonstrate the feasibility of the proposed Structured ACL model in this paper by applying the real SCADA network traffic.**

Keywords*— *Industrial Control System, SCADA, Whitelist, Traffic Locality, Frequent Pattern Mining

I. INTRODUCTION

A control system is a computer based system which is used in various basic facilities and industries to monitor and control whole specific industrial sites or industrial complex. SCADA system is one of the control systems begun to be used in monitoring and controlling remote systems since 1960s. Since SCADA system was operated by private protocol in a closed network, threats related to security were not that great. Through recent integration with business systems, SCADA network has become extended increasing the connection between closed networks, business networks and the entire internet. Cooperation with external partners made the SCADA system more connective and applicable in various media. For such reasons, there have been a tendency for security vulnerabilities to increase. With the gradual liberalization of system by development of technology, the issue of the security in SCADA system is turning from physical security to electronic security violations by hacking, worms, and viruses. Whitelist security technique proves and allow only safe traffic while the Blacklist security technique in contrary blocks malicious traffic. In spite of high security in whitelist, whitelist is only used in restricted areas to avoid serious interferences that it may bring about. But still, it attracts attention as an effective method in ensuring the security in control system environments where there are low resources, regular system operating patterns and network communication traffic.

The SCADA network operates only on a limited basis, with repeated and periodic use of the specified devices. However, the static ACL, which is generally used in the current whitelist security scheme, has a limitation that it does not reflect the mechanical and repetitive communication characteristics of the SCADA system because it contains only simple 5-tuple information.

In this paper, in order to reflect the mechanical and repetitive communication characteristics of the SCADA system, we extract the sequence association relation between the ACL reflecting the two characteristics of periodicity and sequence. The extracted order association relations construct a structured ACL, which is a more specific white list reflecting the mechanical communication characteristics of the SCADA system, in addition to the static ACL.

Section 2 describes related work. Section 3 defines the limitations of existing static ACL and the various ACL that make up the structured ACL proposed in this paper. Section 4 describes the proposed structured ACL model. Section 5 describes the experiment and the results. Section 6 presents the conclusion and future research.

II. RELATED WORK

Static ACL are used in various security products such as network switches, firewalls, and so on. CISCO, Juniper Networks, and McAfee, international information security companies, are applying Static ACL to their network switches and firewalls. In addition, ArubaNetworks, which became a hot topic in the wireless LAN market, is also using Static ACL for wireless LAN switches. Static ACL are used not only in control systems but also in various places. However, since it is very difficult for a person to create a static ACL to be applied to an individual device, various studies on method of automatically generating a static ACL have been conducted [2][6].

Choi seungoh et al[2] suggests that the general characteristics of the SCADA system such as TCP

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2015R1D1A3A01018057) and by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No. 2017-0-00513)

handshaking and common ports are insufficient to generate a flow-based whitelist, and suggest a method to automatically generate a flow whitelist based on the locality of the traffic. In study of Choi seungoh et al[2], whitelist was automatically generated by using locally-frequently-used port and degree centrality only by using 5-tuple information of flow and verified by applying it to actual SCADA system traffic.

However, the static ACL created through the study on the automatic generation of the existing static ACL has a limited expressiveness in expressing the characteristics of the SCADA network. In addition, there are disadvantages such as generation of a lot of rules such as ANY-ANY and generation of rules even if communications occurs only once.

In this paper, we propose a Structured ACL that reflects the features of the SCADA system by adding order associations to the static ACL to overcome the limitations of the static ACL expressions and the disadvantages mentioned above.

III. Static ACL and their limitations

In this section, we define the existing static ACL model and describe three limitations of the static ACL model

A. Static ACL

Static ACL, which generally represent whitelists, are a simple network and system access list based on only simple 5-tuple information and cannot express the order or condition between ACL.

B. Limitations of Static ACL

The first limitation in SCADA systems is the use of only ANY-ANY rule to express the Dynamic Allocated Ports as Static ACL model. But, if you allow the ANY-ANY rule, you must always open all ports for all connections between the IPs. This means that other ACL rules generated between the existing IPs become meaningless. In particular, FTP-enabled servers are likely to be open to anyone accessing that server.

The second limitation is that it does not fully reflect the communication characteristics of the control device. Control devices are devices that need to perform repetitive tasks daily. However, using the static ACL model only expresses the condition that a given task is performed.

The third limitation is that all rules written in the static ACL model, regardless of frequency, are always open.

For example, the management terminal designation for SCADA network equipment management service is used just few times a year, but it must be open all year around.

IV. Proposed Structured ACL model

In this section, we define the proposed structured ACL and its various ACL that make up the structured ACL model, to overcome the three limitations of static ACL. We also describe a method for extracting associations between static ACL and expressing them as structured ACL.

A. Structured ACL Model

Figure 1 shows the concept of creating a structured ACL by extracting the defined associations such as usage time and order information from static ACLs. Each of the six static ACLs on the left has color sequence information, and the order of these colors is extracted through Correlation Finder. The four right structured ACLs have the form of associating static ACLs by extracting the order of purple -yellow, red-blue, and red-green from the left six static ACLs.

In this paper, we define and use a structured ACL that can overcome the three limitations of the static ACL described above and extends the expression power to express the characteristics of the control system traffic.

A structured ACL is a set of ACLs that have a defined order by extracting associations from the control system network traffic. Structured ACL consist of frequent ACL which extracts frequent static ACL, periodic ACL which extracting the periodicity of static ACL, sequential ACL which extracts association between static ACL, and static ACL which were not extracted through three ACL. Static ACL, frequent ACL, periodic ACL and sequential ACL constituting a structured ACL are defined as shown in table 1.

Both frequent ACL and periodic ACL are ACL that extract the period of an ACL. But, analysis of the actual traffic resulted in the problem in which too many sequences that were extracted by ACL are used frequently. We also distinguished between frequent ACL and periodic ACL in order to avoid unnecessary sequences and prevent the case where too frequently occurring ACL are structured as structured ACL.

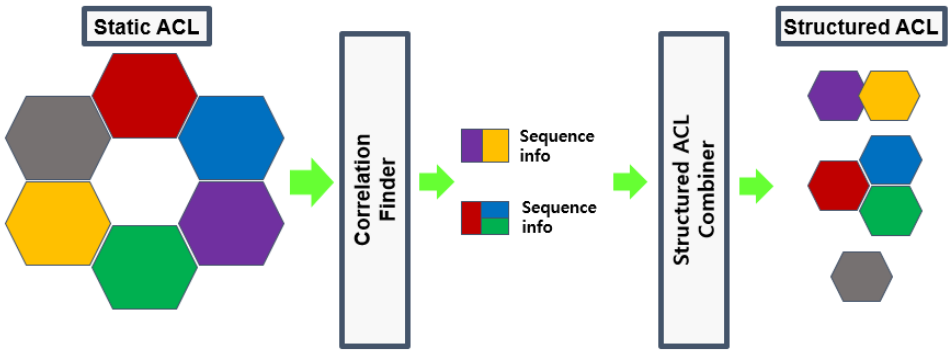

Fig. 1. Concept of Structured ACL

Table I. Definition of ACL

Term	Definition
ACL	A list of users who have set permissions to access specific systems, such as directories or files.
Static ACL	List of access rights to simple network and system based on 5-tuple
Frequent ACL	A list that extracts the repetition frequency of frequent static ACL
Periodic ACL	The extraction of the list repeated cycles of Static ACL
Sequential ACL	A list that extracts association relations between sequences and conditions among static ACLs
Structured ACL	A list of access rights expressing association relations such as order or condition between static ACLs (Structured ACL has static ACL, frequent ACL, periodic ACL and sequential ACL as components.)

Figure 2 shows the relationship between static ACL, frequent ACL, Periodic ACL, Sequential ACL and structured ACL. First, we extract the repeat period of each ACL. Every ACL has a repeat period, so we call them Periodic ACL. But if the repeat period of target ACL is lower than 30 second, then we called them Frequent ACL. Second, we extract the associations between sequences and conditions among Static ACLs, then we called them Sequential ACL. These Periodic ACL, Frequent ACL and Sequential ACL has an inclusion relationship. Finally, Frequent ACL, periodic ACL, sequential ACL and static ACL extracted from existing static ACL constitute structured ACL.

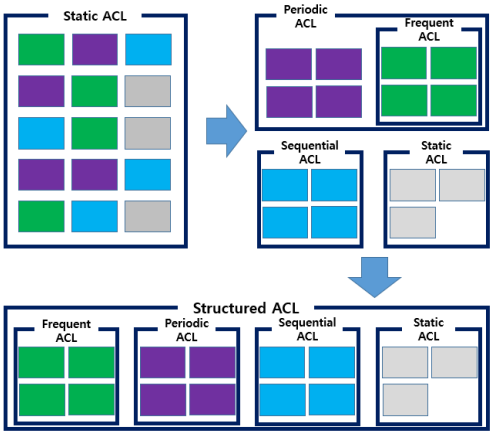

Fig. 2. Relationship between defined ACL

B. *Association Relation Extraction*

The structured ACL proposed in this paper is constructed by adding periodicity and sequence association relations to static ACL. The process of extracting the sequence association relation between static ACL to construct a structured ACL consists of three steps.

In the first step, each static ACL to be used for structured ACL configuration is assigned a unique number. This is used to extract matching information by matching the static ACL to the traffic to generate the transaction thereafter.

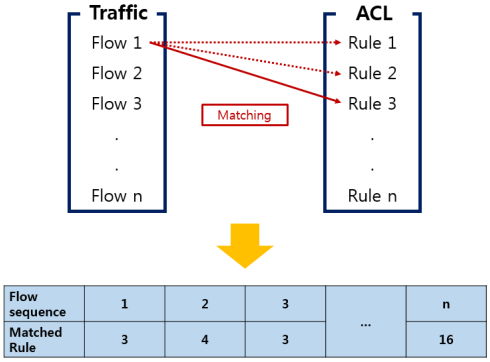

Fig. 3. Process of match ACL to traffic

The second step is the transaction creation phase. In this step, a transaction is generated using static ACL and the traffic is used to generate the static ACL. AS shown in Figure 3, the transaction matches the traffic to the static ACL by the flow unit, and the unique numbers of the matched static ACL is listed in chronological order of the flow. Figure 3 shows the second phase of the association relation extraction. Static ACL is matched to each flow of input traffic and the unique numbers of matched ACL is listed in chronological order.

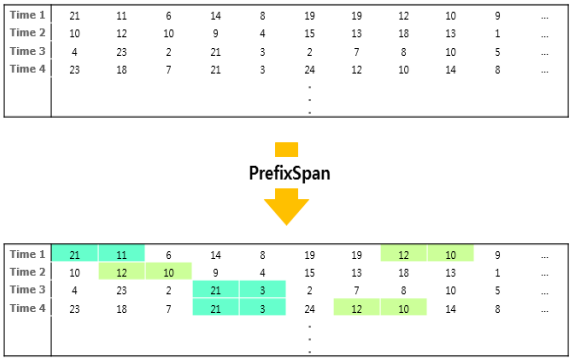

Fig. 4. Concept of sequence correlation extraction

The third step is the sequence information extraction step. In this step, we use the PrefixSpan algorithm, which is a sequential pattern mining algorithm, to extract the association between transaction ACL matching sequences. Figure 4 shows the process of extracting the association between transaction and ACL matching sequence using PrefixSpan algorithm We can see that No. 11 or No. 3 ACL match after No. 21 ACL, and No. 10 ACL match after No. 12 ACL. Figure 5 shows an example of a structured ACL expressing the order relation that No. 12 ACL is followed by No. 10 ACL. The structured ACL consist of independent No. 12 ACL and No. 10 ACL with structured ACL with one hierarchical structured through association extraction.

The structured ACL shown in Figure 5 can reduce the unnecessary detection of the firewall by matching No. 10 ACL only when the No. 12 ACL is matched through the hierarchical structure. Also, since it contains Max duration information of each ACL, when the corresponding ACL is matched, it just opens the rule only for Max duration time, so that the risk of opening all the rules always can be avoided.

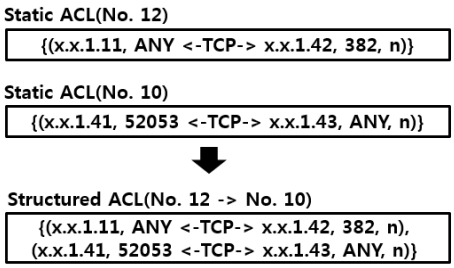

Fig. 5. Example of structured ACL Model

V. EXPERIMENT

In this chapter, we test our proposed structured ACL model to real SCADA network traffic to test its feasibility. Table 2 shows the SCADA network traffic used in the experiments. The remaining 2 sites excluding the A-site are 12-hour traffic collected from each site.

Table II. Test set spec

Site	Size of Traffic	No. of Flows
A	492 GB	9,198,938
B	14.6 GB	234,259
C	3.14 GB	13,729

Table 3 summarizes the results of applying the proposed system for each site. A site has 655 total static ACL, 82 extracted frequent ACL, 27 periodic ACL from 1 minute to 5 minutes, and 30 sequential ACL. The number of ACL elements that exclude duplicated element from sequential ACL was 18, and the number of static ACL that was not extracted is 637.

Table III. Experiment result

Site	Static ACL	Frequent ACL	Periodic ACL	Sequential ACL	Left Static ACL
A	655	82	27	30(18)	637
B	38	4	2	2(2)	36
C	90	11	3	0	90

Table 4 shows the periodic ACL in more detail. A periodic ACL with a period between 1 and 5 minutes was divided by minutes.

Table IV. Periodic ACL in more detail

Site	1-min Periodic ACL	2-min Periodic ACL	3-min Periodic ACL	4-min Periodic ACL	5-min Periodic ACL
A	3	9	7	3	5
B	0	0	0	0	2
C	0	1	0	0	2

As a result, the frequency of 82 ACL among the total 655 static ACL extracted by the A site traffic was extracted as 30 or less and excluded from the structured ACL extraction process. In addition, it was possible to extract repetition period of less than 5 minutes in 27 ACL. The total number of association relations between extracted ACL was 30, and the total number of unique ACL with duplicates removed was 18. Through experiments, it was possible to construct a structured ACL with hierarchical structure by extracting the association relation between 18 ACL out of 655 static ACL.

VI. CONCLUSION AND FUTURE WORK

In this paper, we propose a structured ACL model that adds periodicity and sequence extraction order associations to static ACL. The proposed structured ACL contains various information such as sequence relation between each rule, duration, S / C information, repetition cycle, and overcomes the three limitations of existing static ACL as presented above. We have also proved that the proposed scheme is feasible in actual SCADA traffic.

Future research will include a method to detect abnormalities in the flow by modeling the packet statistical information

REFERENCES

[1] Yun, Jeong-Han, et al. “Burst-based anomaly detection on the DNP3 protocol.” International Journal of Control and Automation 6.2 (2013): 313-324.

[2] Choi, Seungoh, et al. “Traffic-Locality-Based Creation of Flow Whitelists for SCADA Networks.” Critical Infrastructure Protection IX. Springer International Publishing, 2015. 87-102.

[3] Schneider, Johannes, Sebastian Obermeier, and Roman Schlegel. “Cyber security maintenance for SCADA systems.” Proceedings of the 3rd International Symposium for ICS & SCADA Cyber Security Research. British Computer Society, 2015.

[4] Igure, V.M. and Laughter, S.A. and Williams, R.D.: Security issues in SCADA networks. Computers & Security. Vol. 25, no. 7, pp. 498–506, Elsevier (2006)

[5] Jung, Woo-suk, et al. “Whitelist representation for FTP service in SCADA system by using structured ACL model." Network Operations and Management Symposium (APNOMS), 2016 18th Asia-Pacific. IEEE, 2016.

[6] Barbosa, Rafael Ramos Regis, Ramin Sadre, and Aiko Pras. "Flow whitelisting in SCADA networks." International journal of critical infrastructure protection 6.3 (2013): 150-158.

[7] R. Barbosa, R. Sadre and A. Pras, A first look into SCADA network traffic, Proceedings of the IEEE Network Operations and Management Symposium, pp. 518–521, 2012.

[8] Kang, Dong-Ho, et al. "Whitelist Generation Technique for Industrial Firewall in SCADA Networks." Frontier and Innovation in Future Computing and Communications. Springer

[9] Kang, Dong-Ho, Byoung-Koo Kim, and Jung-Chan Na. "Cyber threats and defence approaches in SCADA systems." Advanced Communication Technology (ICACT), 2014 16th International Conference on. IEEE, 2014.

[10] L. Freeman, Centrality in social networks: Conceptual clarification, Social Networks, vol. 1(3), pp. 215–239, 1978-1979.