Improve the Firewall Accuracy By using Dynamic Ontology

Conference Paper · November 2020

DOI: 10.1145/3440749.3442607

Mohammed Hashayka
An-Najah National University
1 PUBLICATION **0** CITATIONS
SEE PROFILE

Amjad Hawash
An-Najah National University
33 PUBLICATIONS **68** CITATIONS
SEE PROFILE

Ahmed Awad
An-Najah National University
31 PUBLICATIONS **63** CITATIONS
SEE PROFILE

Othman Othman
An-Najah National University
17 PUBLICATIONS **28** CITATIONS
SEE PROFILE

Improve the Firewall Accuracy By using Dynamic Ontology

Qossay Ismail
qusayismail9@gmail.com
Department of Networks & Information Security
An-Najah National University
Nablus, Palestine

Osama Saleh
osamasaleh-nis@outlook.com
Department of Networks & Information Security
An-Najah National University
Nablus, Palestine

Mohammed Hashayka
mohammedhashayka@gmail.com
Department of Networks & Information Security
An-Najah National University
Nablus, Palestine

*Ahmed Awad
ahmedawad@najah.edu
Department of Networks & Information Security
An-Najah National University
Nablus, Palestine

*Amjad Hawash
amjad@najah.edu
Department of Networks & Information Security
An-Najah National University
Nablus, Palestine

Othman Othman
othman.omm@najah.edu
Department of Networks & Information Security
An-Najah National University
Nablus, Palestine

Abstract

Data is considered an important asset for organizations, companies, and even people. Crucial decisions depend mainly on data. Exchanging data is essential in order to negotiate ideas, thoughts, and decisions. Networks are the communication channels of data exchange although data is exposed to different attacks, threats, and loss. Because of this, data security has become a key concern for different parties through their daily data manipulation. There are different ways to ensure data security. Paying attention to network threats, data encryption, and using strong passwords are all examples. However, a firewall represents the first defense line against malicious traffic throughout the network. Firewalls have a set of rules to be applied in the time of data exchange between inside and outside of data networks. Some of the firewalls apply such rules in a sequential manner, which degrades the performance of the firewall. In this work, we are utilizing a dynamic ontology of different firewall rules managed by SPARQL queries, so that the rules are applied faster, and thus, increasing the firewall performance. Experimental results show that our proposed methodology totally eliminates the anomalies in the firewall rules as a result of conducting longest matching with proper rules from the dynamically constructed ontology.

CCS Concepts: • **Computer Communication Networks** → **Security and Protection**;

Keywords: Firewall, Ontology, Rule, False Positive, False Negative, Anomalies, Shadowing, Correlation, Redundancy, Generalization, SPARQL, RDF, DOM, Jena, Port, IP, Software Defined Network.

ACM Reference Format:
Qossay Ismail, Osama Saleh, Mohammed Hashayka, *Ahmed Awad, *Amjad Hawash, and Othman Othman. 2020. Improve the Firewall Accuracy By using Dynamic Ontology. In *ACM International Conference on Future Networks & Distributed Systems.*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Firewall [1] is the first way to protect the networks from external attacks or unwanted data. Firewalls work by having rules inside them. When a data packet passes through a firewall, the firewall checks if any of its rules matches the header of that packet, then the firewall takes the action either to allow or deny the packet. Most of the firewalls check the packets against its rules sequentially which leads to a false-positive/negative decision [3] . Some of such problems/anomalies are:

- Shadowing: occurs when all the rules that match a later rule, match a preceding rule.
- Correlation: occurs when two rules with different actions share a set of packets.
- Redundancy: occurs when two rules with the same action match a set of packets. Thus, if one of the rules is removed, the security policy will not be affected.

[*] Corresponding Authors: ahmedawad@najah.edu and amjad@najah.edu.

ICFNDS'20, Nov 26–27, 2020, ST.PETERS-BURG, Russia

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

- Generalization: occurs when a preceding rule matches packets that match a later rule and the rules perform different actions.

Several techniques have been proposed in the literature to detect the anomalies in firewall rules [2]. However, some techniques do not achieve the minimum level of accuracy required for networks wherein critical data is being exchanged. Furthermore, periodically changeable domains such as firewall rules should be modeled in a simplified fashion such that their maintainability is preserved. Finally, the time required to detect an anomaly represents another concern that should be considered especially in real-time networks.

In this work, we propose a methodology to improve the accuracy of a firewall through modeling firewall rules using a dynamic ontology, which has shown great evidence in simplifying periodically changeable data. Once constructed, the SPARQL queries are then used to check the proper action of the firewall. Finally, we confirm the accuracy of the proposed methodology by comparing the action that should be taken in case of an anomaly with the action that is actually taken when the firewall rules are anomaly-free. Our contributions are summarized as follows:

- A methodology is proposed to construct a dynamic ontology for firewall rules.
- Proper SPARQL queries are used to check the action of the firewall upon the arrival of a packet.
- The accuracy of the propped methodology is evaluated in terms of the deny ratio when a set of random traffic is applied on a firewall.

. The rest of this paper is organized as follows: Previous work is proposed in Section 2. Section 3 defines the problem statement while the proposed methodology is discussed in Section 4. Section 5 demonstrates the experimental results. Finally, Section 6 concludes this paper.

2 Related Work

The firewall represents the first defense line against malicious traffic in a network. Static rule mapping has been adopted in the literature to build the set of rules associated with a firewall [12, 13]. However, detecting anomalies in such firewalls is a difficult task. Moreover, firewall rules are continuously changeable in the network in accordance with the changes in the policy of an organization [10]. Therefore, firewall rules should be constructed in an easily maintainable fashion with the objective of minimizing anomalies.

A lot of effort is evident in the literature to detect firewall anomalies [4, 7, 11]. In [2], Al-Shaer et al. defined the anomalies that may exist in a firewall rule set based on the relations between rule fields (protocol, source and destination addresses, port, and action). The authors in [3] have developed a tool to detect anomalies in multi-vendor firewalls. However, the accuracy of such algorithms should be enhanced to keep pace with the dynamic changes in the rules following the policy enhancements. Other anomalies detection and elimination algorithms have been proposed in [5, 8, 9].

In [6], the authors employ the use of semantic web languages (ontology) with SPARQL queries to detect anomalies. However, static ontology has been constructed in this work which degrades the maintainability of the representation when rules have to be manipulated.

Because the firewall rule set is frequently changeable domain, we propose the use of a dynamic ontology to represent firewall rules which enables us to check the accuracy of firewall decisions through employing the proper SPARQL queries.

3 Problem Statement

Given a set of firewall rules, our objective is to build a dynamic ontology representing those rules to minimize the number of false-positive or false-negative actions (anomalies) upon the arrival of a packet. The action is taken as a response to the SPARQL query when applied to the constructed ontology. This, in turn, helps the network administrator in auditing and tracking the firewall decisions.

For a given firewall, the Deny Ratio (DR) is defined as the percentage of the number of denied packets to the whole number packets sent to that firewall in a given window of time, as formulated in eq.(1) where $DP\#$ represents the number of dropped packets and $TP\#$ represents the total number of packets received by the firewall.

$$DR = (\frac{DP\#}{TP\#}) * 100\% \tag{1}$$

To evaluate the accuracy of a firewall, the DR of that firewall is computed and then compared with the DR of the ideal case of that firewall (when all anomalies are eliminated). Based on that, the accuracy of a firewall is formulated in eq.(2) where DR represents the deny ratio of that firewall, and DR_I represents the deny ratio of the anomaly-free version of that firewall.

$$Accuracy = 100 - (\frac{|DR_I - DR|}{100}) \tag{2}$$

It is important to mention that the proposed approach is applied for auditing purposes to discover the existence of anomalies. However, it can be performed as well in a live environment using techniques like Software Defined Network (SDN).

4 Proposed Methodology

To validate the obtained accuracy of a dynamic ontology-based firewall, we compare the accuracy of the classical sequential firewall with an ontology-based firewall. This comparison is conducted in terms of the Deny Ratio (DR).

4.1 Proposed Algorithm Details

Figure 1 illustrates the general phases of our proposed algorithm for comparison which can be summarized as follows:

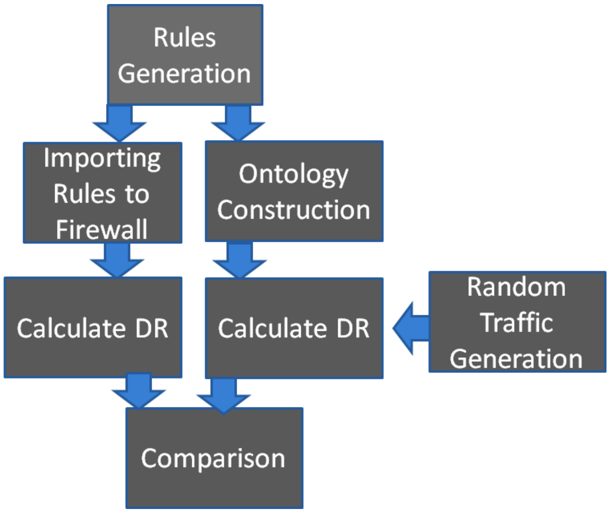

Figure 1. Methodology General Flow.

1. Generation of firewall rules: A set of random rules is generated from a set of random IP addresses and sub-nets. The resultant rules are shuffled as well to appear in different orders so that the task of anomaly detection gets more complicated.
2. Exporting rules to a firewall: The resultant rules from the previous phase are exported to a firewall that behaves in a sequential manner. This is usually done by the administrator either manually or using some automation tool. Typically, this phase is applied periodically in accordance with the changes in the policy of the organization.
3. Ontology construction: The resultant rules from the first phase are converted to XML and Jena library[1] is used then to build the ontology which is represented by Resource Description Language (RDF).
4. Traffic generation: Random traffic is generated using a ping tool with the support of needed scripts to randomize the packet headers. The generated traffic is applied for a predefined time window for both the classical firewall and the ontology-based firewall. Notice that each copy of the packet is parsed as a SPARQL query when sent to the ontology-based firewall so that the RDF file is searched and the proper action is returned.
5. Comparison: The Deny Ratio (DR) within the predefined time window is computed for both the sequential firewall and ontology-based firewall. Thereafter, the accuracy is computed following eq.(2) for both firewalls when compared with the ideal case.

Figure 2 illustrates the detailed flowchart of the proposed algorithm. Notice that in this paper pfSense[2]firewall has been used. However, this algorithm can be used in any firewall regardless to its vendor.

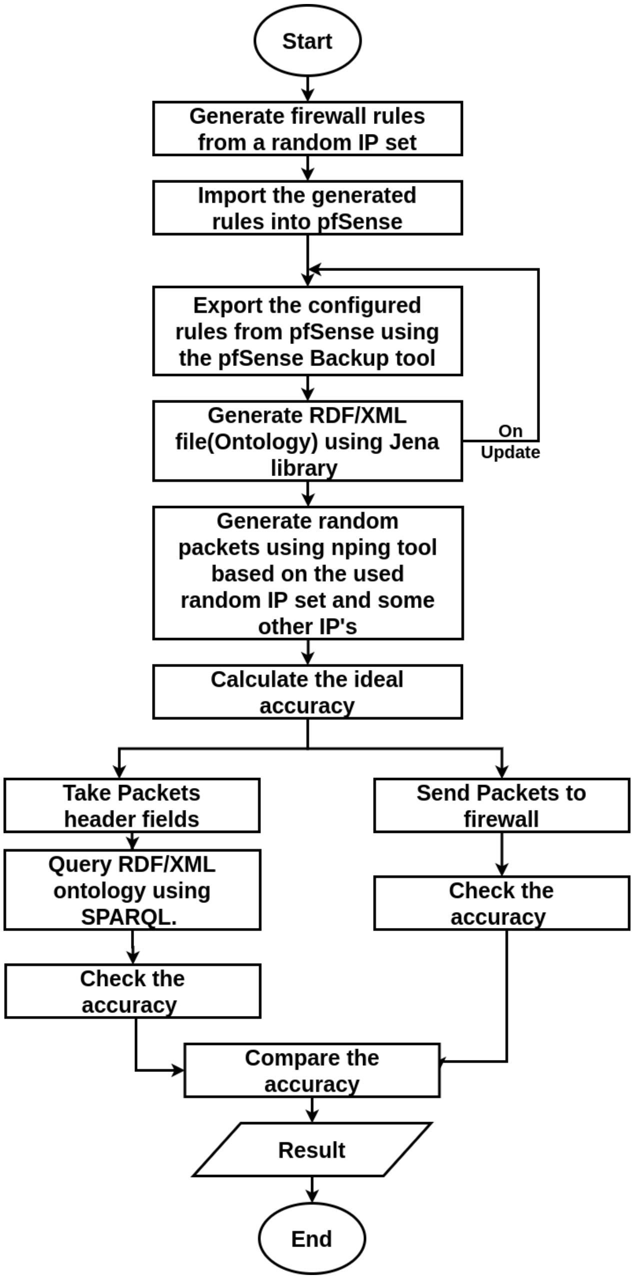
Start
Generate firewall rules from a random IP set
Import the generated rules into pfSense
Export the configured rules from pfSense using the pfSense Backup tool
Generate RDF/XML file(Ontology) using Jena library
On Update
Generate random packets using nping tool based on the used random IP set and some other IP's
Calculate the ideal accuracy
Take Packets header fields
Query RDF/XML ontology using SPARQL.
Check the accuracy
Send Packets to firewall
Check the accuracy
Compare the accuracy
Result
End

Figure 2. Detailed Algorithm.

[1] https://jena.apache.org

[2] https://www.pfsense.org/

4.2 Time Complexity Analysis

The performance of firewalls during the decision making process is a crucial parameter. It is essential that a firewall executes the filtering process in a short time in order to avoid additional congestion burden on the network. Of course, applying filtering rules in a sequential way decreases a firewall performance and hence increases the congestion. On the other side, applying ontological-based filtering rules includes searching for these rules by applying special SPARQL query that includes a fast ontology (graph) traversal algorithm. This leads to a direct[3] and thus, faster action to be conducted by the firewall. Although we do not use an optimized SPARQL query in this work and we do not compare the performance between applying sequential and ontological-based filtering rules, our main contribution focuses on comparing the accuracy of filtering rules between the sequential and ontological-based rules.

Moreover, filtering rules sequence (the way the rules are stored in a queue of rules) plays a dominant role in the decision conducted by the sequential-based firewall. For example, if a firewall contains 100 filtering rules, and one of those rules is positioned at the end of the queue and, suppose there is a huge demand for this rule during the online filtering process of packets. Definitely, the firewall performance will degrade tremendously. Therefore, in this case, positioning this rule at the beginning of the queue is crucial. This needs some AI technique to keep sorting the rules inside the queue (priority Queue in this case) to put the most applied rules at the beginning of the queue as a try to minimize their sequential access times. Whereas, if those rules are constructed in an ontological fashion, taking into account their relations, this leads to a flattering ontology with a graph radius that is much less than the 100 sequential sequences of rules. In this case, there will be no need to sort the rules according to their executing frequency since the access to those rules is executed with a few ontology traversal steps.

5 Experimental Results

In our experiments, we used 3000 packets as sample to check the accuracy of both the sequential firewall and ontology-based firewal, we used nping tool to generate this stream of packets. We have constructed 5 test cases where in each case a specific number of IPs and ports are used to build the rules in a random manner with proper shuffling for the constructed rules. Java code is written to build the rules in each case.

The generated packets in each test case are applied to both sequential firewall and ontology-based firewall. In addition, the ideal Deny Ratio (DR) (in case of anomaly-free firewall) is computed for each test case. The obtained results are shown in Table 1, wherein the average DR for each firewall is illustrated. Notice that in case of sequential firewall, we have recorded the minimum and maximum DR as well because for each shuffle, the decision in this firewall is affected. However, for the ontology-based firewall, the decision is made directly as a response to the proper SPARQL query with the longest matching rule. Therefore, the DR is the same regardless to the shuffle process.

The DR of the ontology-based firewall has been found the same as for the ideal case of the firewall. This means 100% accuracy which demonstrates the effectiveness of such approach if compared with the sequential firewall whose average accuracy degrades when the number of rules increases. This result is reasonable as the probability of anomalies increases when more rules are applied to the firewall, which turns out into some wrong actions in the sequential firewall. However, in case of the ontology-based firewall, our efficient SPARQL query results in fetching the longest match rule from the ontology, and thus, overcoming the anomalies with preserving 100% accuracy in the firewall.

Figure 3 illustrate the DR for both the ideal firewall (and the ontology firewall) and the sequential firewall. Notice that the difference between the curves clearly increases when more rules are applied to the firewall. This is reasonable as the probability of anomalies increases as well which results in wrong actions conduced by the sequential firewall.

6 Conclusions & Future Work

Firewall decisions may have a lot of false-positive/negative actions when the number of rules is increasing. This is reasonable as a high number of rules may result in different types of anomalies existence. Therefore, we have proposed an approach that can eliminate such anomalies based on dynamic ontology. Experimental results have shown great accuracy if compared with the classical sequential firewall.

Our future work will focus on checking the performance of taking firewall actions based on dynamic ontology in terms of the run-time, and the application of dynamic ontology and SPARQL queries in Software Defined Network (SDN).

References

[1] Habtamu Abie. 2000. An Overview of Firewall Technologies. 96 (12 2000), 47–52. Issue 3.

[2] Ehab S. Al-Shaer and Hazem H. Hamed. 2003. *Firewall Policy Advisor for Anomaly Discovery and Rule Editing.* Springer US, Boston, MA, 17–30. https://doi.org/10.1007/978-0-387-35674-7_2

[3] E. S. Al-Shaer and H. H. Hamed. 2004. Discovery of policy anomalies in distributed firewalls. In *IEEE INFOCOM 2004*, Vol. 4. 2605–2616 vol.4.

[4] G. Betarte, E. Gimenez, R. Martinez, and A. Pardo. 2018. Improving Web Application Firewalls through Anomaly Detection. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. 779–784. https://doi.org/10.1109/ICMLA.2018.00124

[5] C. Chao. 2011. A flexible and feasible anomaly diagnosis system for Internet firewall rules. In *2011 13th Asia-Pacific Network Operations and Management Symposium*. 1–8.

[6] R. F. Cordova, A. L. Marcovich, and C. A. Santivanez. 2018. An Efficient Method for Ontology-Based Multi-Vendor Firewall Misconfiguration

[3] Of course this depends on the Ontology size.

# IPs	# of Rules	Ideal DR	Sequential Firewall				Ontology		Better?
			Min DR	Max DR	Avg DR	Accuracy	DR	Accuracy	
5	83	0.935	0.935	0.938	0.936	(99.9 , 99.9, 99.9)%	0.935	100%	Ontology
20	336	0.753	0.728	0.755	0.740	(96.6 , 99.9, 98.2)%	0.753	100%	Ontology
50	455	0.837	0.810	0.841	0.826	(96.7 , 99.9, 98.6)%	0.837	100%	Ontology
80	485	0.789	0.742	0.789	0.766	(94.0 , 99.9, 97.1)%	0.789	100%	Ontology
150	918	0.748	0.638	0.719	0.677	(85.5 , 96.1, 90.5)%	0.748	100%	Ontology

Table 1. Sequential Vs. Ontological-Based Firewall Rules Comparison.

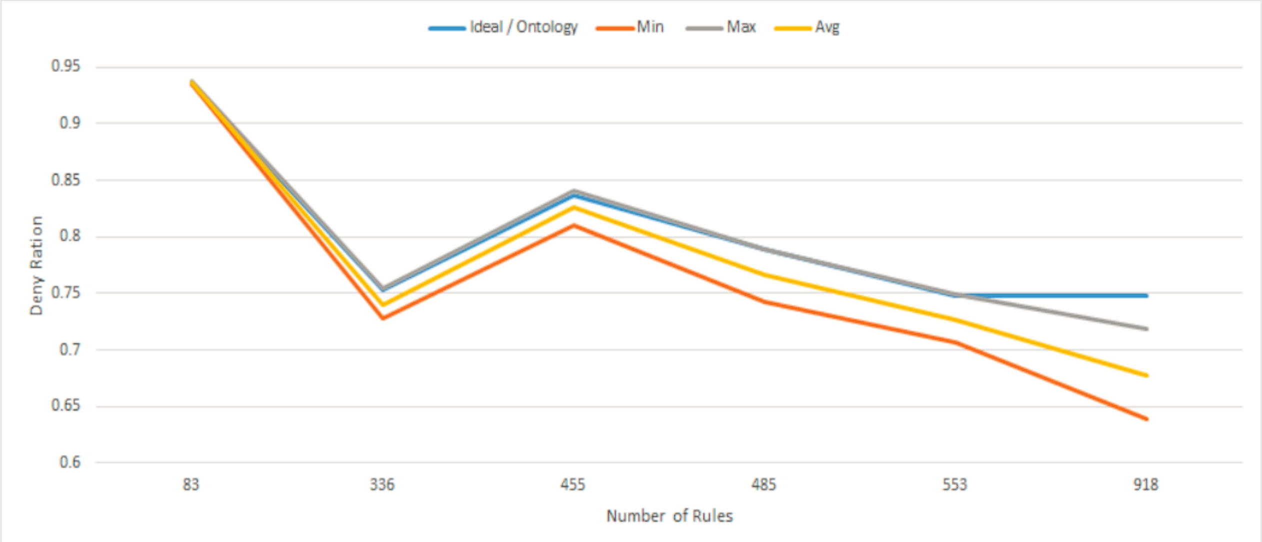

Figure 3. Deny Ratio for Different Number of Rules.

Detection: A Real-Case Study. In *2018 IEEE ANDESCON*. 1–3.

[7] D. Dmitry, P. Elena, C. Anna, Z. Tatiana, and P. Elena. 2020. Approaches to Anomaly Detection in Web Application Intrusion Detection Systems. In *2020 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*. 532–535. https://doi.org/10.1109/USBEREIT48449.2020.9117745

[8] S. Khummanee, A. Khumseela, and S. Puangpronpitag. 2013. Towards a new design of firewall: Anomaly elimination and fast verifying of firewall rules. In *The 2013 10th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. 93–98.

[9] K. Lubna, R. Cyiac, and Kavitha Karun A. 2013. Firewall log analysis and dynamic rule re-ordering in firewall policy anomaly management framework. In *2013 International Conference on Green Computing, Communication and Conservation of Energy (ICGCE)*. 853–856.

[10] Ratish Mohan, Anis Yazidi, Boning Feng, and B. John Oommen. 2016. Dynamic Ordering of Firewall Rules Using a Novel Swapping Window-Based Paradigm. In *Proceedings of the 6th International Conference on Communication and Network Security* (Singapore, Singapore) *(ICCNS '16)*. Association for Computing Machinery, New York, NY, USA, 11–20. https://doi.org/10.1145/3017971.3017975

[11] A. Saâdaoui, N. B. Y. B. Souayeh, and A. Bouhoula. 2015. Automated and Optimized FDD-Based Method to Fix Firewall Misconfigurations. In *2015 IEEE 14th International Symposium on Network Computing and Applications*. 63–67. https://doi.org/10.1109/NCA.2015.31

[12] R. K. Sharma, H. K. Kalita, and B. Issac. 2014. Different firewall techniques: A survey. In *Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*. 1–6.

[13] N. Zope, S. Pawar, and Z. Saquib. 2016. Firewall and load balancing as an application of SDN. In *2016 Conference on Advances in Signal Processing (CASP)*. 354–359.