DoS Detection on SDN Architectures Using Parametric Statistical Tests

PAULO PINTO[1], LUIS PINTO[2]
[1] Novabase IMS, Lisboa, PORTUGAL
paulo.pinto@novabase.pt, www.novabase.pt
[2] UBI university and CITAD, Covilhã, PORTUGAL
moreirapinto.arq@clix.pt, www.ubi.pt

Abstract: - The work realized in this paper as a twofold approach: by one side to present a proof-of-concept that illustrate the flexibility and ease of anomaly detection algorithm implementation on a SDN architecture accessing to the information (counters) existing on the controller; by other side to show the applicability of nonparametric statistics, Kendall's W statistic, to detect network changes on SDN corresponding to possible DoS attacks.

Key-Words: - Architecture; Software Defined Networks; Denial of Service; Kendall; Parametric Statistics

1 Introduction

Software Defined Networks (SDN), an operational and programming network architecture, facilitates new opportunities for innovative security applications. The SDN controller, acting as an interface point to the network, aggregates information flows that can be used by security management applications to gather feedback from network as to their well-being states. While information flow processing and analysis is a mature field its application is only beginning in SDN and there are many open questions on its use.

This article proposes an algorithm that uses hypotheses testing with Kendall's W (coefficient of concordance) statistics to identify significantly associated, or concordant, groups of flows in a SDN indicating possible network attacks that are difficult to prevent by standard settings of information resource software, namely "Denial-of Service Attacks (DoS)".

Our work focuses on the statistical variation analyses of the traffic flows, which is a performance measure of particular interest for network engineering (i.e., provisioning, SLA definition, anomaly detection, etc.). The innovative aspect of our work is the use of nonparametric statistical test to detect a variation on it the network state. The proposed statistical test does not assume an underlying model or an underlying model validation. Also, due to a low implementation complexity, O(n), the proposed algorithm can be used to monitor and test data at wired speed links.

2 Software Defined Networks

The last decade has seen an extraordinary revolution in end use equipment. Our servers have grown a thousand-fold in their intelligence and computing capacity. The links have also undergone similar transformation.

The enormous increase in computer power together with the vastly increase connection speed between network elements has made the concept of SDN technologically feasible and economically viable. Software Defined Networking is a new approach to IP networking that centralizes the management and control plane logic in an SDN controller and creates a well-defined interface to the forwarding plane via a standardized protocol (Openflow). The intention of SDN is to spur innovation in service creation and deployment. The network is considered „programmable" and new applications can be easily added by a variety of players.

With SDN the data in the network (e.g. stats, service state, security, etc.) can be analyzed and used by an application to create policy intent and program the network into a new configuration. Programmability (i.e. the ability to access the network via APIs and open interfaces) is central to SDN.

What's key here is that the controller's APIs will deliver a whole of information that paves the way to a new approach on the analytics area providing more wide available and accurate information on the traffic. It could be used in various applications such as detection of anomalies (e.g., denial of service attacks or link failures), prediction of traffic growth, or assessment of the impact on network traffic of a new customer or of a new application.

2 Coefficient of Concordance W

Proposed by Maurice G. Kendall and Bernard Babington Smith, Kendall's coefficient of concordance (W) is a measure of the agreement among several (k) quantitative or semi quantitative variables that are assessing (ranking) a set of N objects of interest. Depending on the application field, the "judges" can be variables, characters, and so on. They are dimensions of communication flows in the present article.

The degree of agreement among k judges is reflected by the degree of variance among the N sums of ranks. W, Kendall's coefficient of concordance, is a function of that degree of variance.

To compute W [6], we first find the sum of ranks, Rj, in each column of a k x N table. Then we sum the Rj and divide that sum by N to obtain the mean value of Rj. Each of the Rj may then be expressed as a deviation from the mean value. The larger the deviations, the greater is the degree of association among the k sets of ranks. Finally, S, the sum of squares of these deviations, is found. Knowing these values, we may compute the value of W:

Equation (1) Kendall's Coeficient of Concordance

$$W = \frac{12\, S}{k^2(N^3 - N)}$$

When N is larger than 7, the expression given by k(N-1)W is approximately distributed as a chi square with df = N-1. That is the probability associated with the occurrence under H0 of any value as large as an observed W may be determined by finding $\chi 2$ = k(N-1)W and then determining the probability associated with so large value of $\chi 2$ referring to its distribution table of values.

3 Algorithm Development

We start by choosing the variables that best characterize a network state in a SDN- traffic models shows that two parameters, the load of the channel and the number of active flows in it, must be used for a full representation of the network state [1]; To instantiate these two parameters we then propose the use of existing counter variables in the SDN controller: "received bytes per flow" and "flows durations" to characterize the load of the channel; and to use "active entries per table" to characterize the number of active flows.

Next, our intention is now to detect network state changes, which were originated by a Denial of Service attack, using this variables and Kendall's W statistic hypothesis tests. Considering that [3] under a simulated attack, is expected an increase in the number of flows, a uniform size of the flows (near 50 bytes) and long streams flows (more than 5 minutes) we can test the variation of this variables on time looking for this kind of change using the Kendall's W statistic hypothesis tests. If a change occur we may expect to find some of these variables associated indicating a possible attack.

These data variables are appropriate to be used with Kendall's W statistics because they are all indirect measures of a common representation of the network state [1].

We now proceeded by generating two appropriately simulated random data tables containing information flows that instantiates the variables chosen in the previous phase and

ISBN: 978-960-474-380-3

characterize a network state: Table I under normal conditions, where is expected a random number of flows, flows size and durations; Table II under a simulated attack, where is expected an increase in the number of flows, a uniform size of the flows (for test proposes we've consider values near 50 bytes) and long streams flows (more than 5 minutes). The data simulating these conditions is presented in Table II, rows 16, 17 and 18.

Considering the hypothesis H0 to be used by Kendall's W Statistics: "There is no association (concordance) among the variables that characterize the network state"; we now proceed to test this hypothesis with the two different datasets contained in Table I, Table II;

Conduct an overall test of concordance of all random flows in the Table I. If not under an attack it is expected no association among the traffic flows, i.e. using data in Table I we expect to accept H0.

Conduct an overall test of concordance of all flows in Table II. If under a DoS attack it is expected that there is at least one variable that is associated with one or some of the others, i.e. using data in Table II we expect to reject H0

TABLE I. RANDOM FLOWS

COUNTER.**PERTABLE**		COUNTER.**PERFLOW**		COUNTER.**PERFLOW**		R_j
active entries		**received bytes**		**duration**		
data	rank	data	rank	data	rank	
15	14	22	7	116	11	32
129	3	18	8	55	15	26
73	9	24	5	102	13	27
149	1	12	10	113	12	23
93	7	25	4	81	14	25
118	5	5	12	311	3	20
23	13	12	10	311	3	26
38	12	29	3	180	9	24
7	15	5	12	205	6	33
51	11	37	1	332	1	13
81	8	3	14	193	8	30
103	6	17	9	319	2	17
142	2	36	2	196	7	11
126	4	1	15	148	10	29
61	10	23	6	224	5	21
15	14	22	7	116	11	32
129	3	18	8	55	15	26
73	9	24	5	102	13	27

For the values in Table I we calculate W getting 0,23 and a correspondent value of k(N-1) as being 9,81 > 23,68 (the value of $\chi 2$ = 23,68 with α= 0,05 and df =14) taking us to accept H0 and conclude that there is no association among the traffic flows.

TABLE II. RANDOM AND SIMULATED ATTACK FLOWS

COUNTER.**PERTABLE**		COUNTER.**PERFLOW**		COUNTER.**PERFLOW**		R_j
active entries		**received bytes**		**duration**		
data	rank	data	rank	data	rank	
101	11	5	17	75	12	40
108	10	32	8	33	16	34
43	16	17	11	26	17	44
24	17	2	18	253	5	40
129	6	31	9	99	11	26
148	4	35	6	241	6	16
122	8	9	14	1	18	40
133	5	10	13	153	9	27
50	14	12	12	326	1	27
71	13	6	15	192	7	35
8	18	34	7	102	10	35
46	15	37	5	55	13	33
97	12	27	10	34	14	36
111	9	6	15	188	8	32
125	7	38	4	34	14	25
200	3	50	3	310	3	9
250	2	55	1	290	4	7
300	1	53	2	320	2	5

For the values in Table II we calculate W getting 0,55 and a correspondent value of k(N-1) as being 28,24 > 27,59 (the value of $\chi 2$ = 27,59 with α= 0,05 and df =17) taking us to reject H0 and concluding that there is at least one variable that is associated with one or some of the others indicating a possible attack.

4 Conclusion

With this work and the proposed algorithm we reach two goals: by one side we have present a proof-of-concept that illustrate the flexibility and ease of anomaly detection algorithm implementation on a SDN architecture accessing to the information (counters) existing on the controller; by other side we have shown the applicability of nonparametric statistics, Kendall's W statistic, to detect network changes on SDN.

As a result, considering the load of the channel and the number of active flows as variables to characterize a network state in a SDN, we observed with the present work that the Kendall coefficient of concordance can be used with success to assess the degree to which a group of network communications flows in a SDN were associated.
For the proof of concept realized on this work, we used a simple random generated data. Our future work consists in an evaluation of the proposed algorithm collecting data in real time on a university campus network.

References:
[1] Chadi Barakat, Patrick Thiran, Gianluca Iannaccone, Christophe Diot, and Philippe Owezarski, Modelling Internet Backbone Traffic at the Flow Level, *IEEE Transactions on Signal Processing*, Vol. 51, Nº. 8, August 2003
[2] F. Afanasiev, A. Petrov, V.Grachev, A. Sukhov, Flow-based analysis of Internet traffic, *Russian Edition of Network Computing*, 5(98), 2003, pp.92-95
[3] Aleksey A. Galtsev and Andrei M. Sukhov, Network attack detection at flow level, *Smart Spaces and Next Generation Wired/Wireless Networking*, Lecture Notes in Computer Science Volume 6869, 2011, pp 326-334
[4] Pierre Legendre, Coefficient of Concordance, *Encyclopedia of Research Design*, Vol. 1. ed. SAGE Publications, Inc., 2005, pp. 164-169
[5] Thomas Nadeau, Ken Gray, *Software Defined Networks*, O'Reilly, 2013
[6] Sidney Siegel, *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill, 1956

ISBN: 978-960-474-380-3