An Analyzer of Computer Network Logs Based on Paraconsistent Logic

Avelino Palma Pimenta Jr., Jair Minoro Abe, Cristina De Oliveira

An analyzer of computer network logs based on Paraconsistent Logic

Avelino Palma Pimenta Junior, Jair Minoro Abe, Cristina Corrêa de Oliveira

Paulista University, Graduate Program in Production Engineering
R. Dr. Bacelar 1212, 04026-002 São Paulo, Brazil

appimenta@gmail.com, jairabe@uol.com.br,
crisolive@ig.com.br

Abstract. In recent years, the network vulnerability events draw the attention to the issue of the information management on the World Wide Web. The detected vulnerability was not only restricted to individuals, but also to enterprises and governments. Over the past decade, networks have become an affordable way for several computer services, but also a major challenge for network managers to maintain its operation. The main problem is the difficulty to deal with big amount of data generated by user requests, which in turn ultimately generate increasing information logs. Moreover, the dynamics of the services can lead to detect false positive and negative ones, so uncertainty is a theme to be considered. The employment of classical logic may not be adequate to solve problems of this nature. The aim of this paper is to present the development of a Paraconsistent analyzer, in order to extract some computer networks patterns of interest.

Keywords: paraconsistent logic, computer networks, pattern recognition, decision-making

1 Introduction

The computer networks currently constitute as the main form of transmitting data and services. Therefore, the task of monitoring the information has turn to be a key factor in technology sectors [1]. The information security issues have existed around since it has been created. However, as the technology goes further and information management systems become increasingly powerful, the issue of information security becomes also increasingly critical [2].

Considering its intrinsic nature, the network operation analysis is based on stochastic events. The argument for this type of methodology is based on the principle that human actions behave as random elements [3]. In fact, the variability of available services is considerable, and therefore the types of user behavior eventually follow this trend.

Some important elements should be considered in data traffic management, such as trustfulness, confidentiality, integrity and reliability [4] [5].

Among the mentioned elements, reliability is the main object of analysis of this article. It can be defined as the capacity to provide access to information systems as soon as they are requested [4]. A system with low reliability ultimately leads to dissatisfaction and low user productivity.

The establishment of a set of criteria should be done to avoid false positives [6], which in turn may even lead to problems of a legal nature. For instance, a significant loss of network data packets can either be interpreted as a malicious attack, as may represent an intense use of the computer network.

It is possible to gather information from network logs of the data packets that pass through the network devices. Data extraction can provide the manager an important tool in decision making.

Some data may be considered interesting to the analysis of the packet traffic, among which are: the origin logical IP address, request time, response waiting time, type of obtained result, the amount of response data in the transaction and the destination logical IP address [7].

Due the stochastic behavior of the networks, the analysis methods based on classical logic may not be a suitable tool for this scenario [8]. A new logical system is needed to deal with it. Therefore, the Paraconsistent annotated evidential logic $E\tau$ has a structure that becomes a natural technique to look for evidence of problems, whether caused both by the standard operation of the network or intentional elements [9]. In the latter case, it may be constituted by users or malicious application [10].

Once again, the use of Paraconsistent logic $E\tau$ arises as a feasible alternative to take decisions under uncertainty, inconsistency and contradiction, in several areas such as robotics, electronics, traffic control, among others [11].

2 Methodology

The development of the proposal is based on the analysis of network data communication over five days and three ranges (mornings, afternoons and evenings), of five hours each. For each range, several parameters were obtained, among which: date and time of the request, the source IP address, destination IP address, type of connection made, the result of the request operation, response waiting time, amount of data response and total transactions.

From the network requests log, it was possible to extract network usage information expressed in Table 1:

Table 1. Network parameters obtained from transactions logs

Day of week	Range	Events	Total transactions	Avarage response time (ms)	*Standard deviation of avarage response time (ms)*	Avarage packet size (bytes)
Monday	8:00 - 12:59	1 até 76157	76157	15433.17534	118597.16	49972.8692
	13:00 - 17:59	76158 até 135333	59175	15349.66649	123864.8334	16631.58448
	18:00 - 22:59	135334 até 193521	58187	25168.16961	179844.285	22746.42033
Tuesday	8:00 - 12:59	1 até 44070	44070	18834.32151	123962.0828	35533.25305
	13:00 - 17:59	44071 até 112514	68443	12579.59023	86841.53468	20589.08378
	18:00 - 22:59	112515 até 148376	35861	24218.55156	117614.5985	28749.04528
Wednesday	8:00 - 12:59	1 até 53900	53900	14365.10788	102525.2609	31278.15891
	13:00 - 17:59	53903 até 108968	55065	16172.74776	158488.8088	44461.68029
	18:00 - 22:59	108969 até 133015	24046	29514.48547	246460.8734	26382.09061
Thursday	8:00 - 12:59	1 até 52319	52319	19118.12858	110954.54	52906.31
	13:00 - 17:59	52320 até 159662	107342	10186.07854	89632.23525	14449.85665
	18:00 - 22:59	159663 até 196237	36574	25835.7653	272687.2837	27913.28298
Friday	8:00 - 12:59	1 até 37178	37178	17740.07359	88674.67718	37964.06786
	13:00 - 17:59	37179 até 143238	106059	9967.793372	122313.1212	19712.07777
	18:00 - 22:59	143239 até 199849	56610	16821.41266	217288.195	13163.5317

Some significant information can be obtained considering the parameter "Standard Deviation" in association with “Average Response Time” as a measure of dispersion and "Average Packet Size". In this case, it is possible to make an association between the lowest standard deviation (86841.53 ms), its average response time (12579.59 ms) and average packets size (20589.08 bytes), which leads to believe that in the period from 13:00 to 17:59 on Tuesday presented the network operating normally, with low response time, even though with a considerable amount of data in transit. On Wednesday, from 18:00 to 22:59, the network had its worst performance, having obtained the largest delay in average response time (29514.48 ms) and slightly higher average packets size compared to the previous example (26382.09 bytes), with a standard deviation slightly below the maximum limit obtained (246460.67 ms). In this case, it may be viable to conclude that the network had dealt with operations problems.

However, during the computer network operation, handle dynamic and highly stochastic events may be a high complexity task. Therefore, a logical analyzer – Para-analyzer [12] will be used upon the data obtained to make an analysis under the light of an artificial intelligence tool. Four parameters shall be used as factors: average response time (R), its standard deviation (D), average packets size (P) and the total transactions (T).

The number of intervals that were selected for each parameter is based on the occurrence of significant variances in the evaluations of favorable and unfavorable evidences by the specialists. A larger number of intervals often presented very close or even repeated values, which in turn would generate unnecessary redundancy in this study.

It is considered that a low response time is a good indicator because it suggests that the network did not suffer consequences of a possible congestion and was able to answer its requests in an acceptable time. For this, three intervals shall be considered, based on the minimum and maximum values obtained from the network log: R1, R2 and R3.

A low standard deviation of the average response time also leads to the belief of a homogeneous network operation. In other words, no significant discrepancies between the hosts in operation were detected. Along with the previous factor, three intervals shall be considered: D1, D2 and D3.

The average packet size is also an important factor, but it has an element of uncertainty that must be considered. Networks with low average size packets may indicate little use, which can be considered a plus. Moreover, networks that suffer attacks should also have this tendency, since the data packets used for this purpose are individually small. Four intervals will be considered: P1, P2, P3 and P4.

Finally, the number of transactions may be considered a significant factor since a high value may suggest problems relating to malicious attacks or high degree of utilization of the network. Once again, four intervals shall be used: T1, T2, T3, and T4.

The concepts of Paraconsistent logic $E\tau$ will be used from this point. According to Abe[13]: "The atomic formulas of the logic $E\tau$ are of the type $p(\mu, \lambda)$, where $(\mu, \lambda) \in [0, 1]^2$ and $[0, 1]$ is the real unitary interval (p denotes a propositional variable)". Therefore, $p(\mu, \lambda)$ can be intuitively read: "It is assumed that p's favorable evidence is μ and contrary evidence is λ.". This will lead to the following conclusion:

- $p_{(1.0, 0.0)}$ can be read as a true proposition,
- $p_{(0.0, 1.0)}$ as false,
- $p_{(1.0, 1.0)}$ as inconsistent,
- $p_{(0.0, 0.0)}$ as paracomplete, and
- $p_{(0.5, 0.5)}$ as an indefinite proposition.

To determine the uncertainty and certainty degrees, the formulas are[10]:

- Uncertainty degree: $G_{un}(\mu, \lambda) = \mu + \lambda - 1$ $(0 \leq \mu, \lambda \leq 1)$;
- Certainty degree: $G_{ce}(\mu, \lambda) = \mu - \lambda$ $(0 \leq \mu, \lambda \leq 1)$;

An order relation is defined on $[0, 1]^2$: $(\mu_1, \lambda_1) \leq (\mu_2, \lambda_2) \Leftrightarrow \mu_1 \leq \mu_2$ and $\lambda_1 \leq \lambda_2$, constituting a lattice that will be symbolized by τ.

With the uncertainty and certainty degrees, it is possible to manage the following 12 output states, showed in the Table 2.

Table 2. Extreme and Non-extreme states

Extreme States	Symbol	Non-extreme states	Symbol
True	V	Quasi-true tending to Inconsistent	QV→T
False	F	Quasi-true tending to Paracomplete	QV→⊥
Inconsistent	T	Quasi-false tending to Inconsistent	QF→T
Paracomplete	⊥	Quasi-false tending to Paracomplete	QF→⊥
		Quasi-inconsistent tending to True	QT→V
		Quasi-inconsistent tending to False	QT→F
		Quasi-paracomplete tending to True	Q⊥→V
		Quasi-paracomplete tending to False	Q⊥→F

All states are represented in Figure 1:

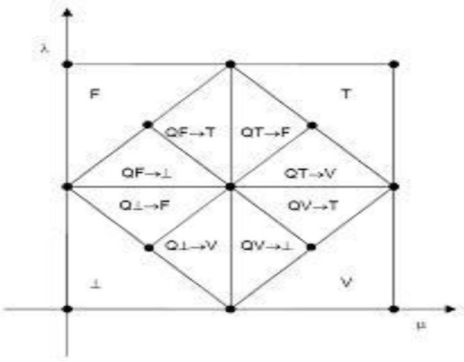

FIG. 1. All states in Lattice τ

Initially, for each analyzed factor, the opinions of two experts in the field of networks shall be considered, both senior professional with a large experience in the field. For each factor, intervals will be taken and rated, with a certain degree of favorable evidence (represented by μ) and unfavorable evidence (represented by λ).

Also weights to each factor/intervals will be applied, considering the importance degree that each expert deems appropriate. The data from which the Paraconsistent algorithm will be applied is applied can be expressed in Table 3.

Table 3. Distribution of factors and grades for the Para-analyzer algorithm

Factor	Interval	Values	Senior Specialist 1		Senior Specialist 2	
			μ	λ	μ	λ
Response Time	R1	< 16666 ms	0.85	0.1	0.9	0.1
	R2	16667 - 23332 ms	0.65	0.45	0.7	0.4
	R3	> 23333 ms	0.45	0.65	0.55	0.75
Standard Deviation of the Average Response Time	D1	< 147999 ms	0.9	0.1	0.9	0.1
	D2	148000 - 209999 ms	0.55	0.5	0.55	0.45
	D3	> 210000	0.2	0.8	0.3	0.8
Average Packets Size	P1	< 19999 bytes	0.7	0.3	0.75	0.3
	P2	20000 – 29999 bytes	0.6	0.4	0.65	0.45
	P3	30000 - 39999 bytes	0.5	0.6	0.55	0.55
	P4	> 40000 bytes	0.2	0.9	0.3	0.85
Transactions	T1	< 39999	0.9	0.2	0.9	0.25
	T2	40000 - 59999	0.7	0.35	0.8	0.3
	T3	60000 - 79000	0.55	0.5	0.6	0.45
	T4	> 80000	0.3	0.8	0.25	0.8

To study the proposition: "The computer network is functioning within normal operational limits", values were tabulated and applied for the Para-analyzer algorithm, as seen in Table 4:

Table 4. Favorable and unfavorable evidences and weights of first scenario

Factor analysis	Interval	Weight	Favorable Evidence Degree	Unfavorable Evidence Degree
Response Time	R1	2	0.9	0.1
	R2	2	0.7	0.45
	R3	2	0.55	0.75
Standard Deviation of the Average Response Time	D1	3	0.9	0.1
	D2	3	0.55	0.5
	D3	3	0.3	0.8
Average Packets Size	P1	1	0.75	0.3
	P2	1	0.65	0.45
	P3	1	0.55	0.6
	P4	1	0.3	0.9
Transactions	T1	2	0.9	0.25
	T2	2	0.8	0.35
	T3	2	0.6	0.5
	T4	2	0.3	0.8

The factors listed above are not able to lead to important conclusions alone. In this case, the combined influence of the factors, with their respective applied weights, could contribute to a more appropriate response to the initial proposition. This is determined by the global analysis of the points that represent the Cartesian plane [14].

The global analysis is calculated considering the favorable evidences (μ) multiplied by their respective weights, and finally added. The same is done to the unfavorable evidence (λ)[14]. Considering the tabulated values, the global analysis obtained was 0.63 of favorable evidence and 0.48 of unfavorable evidence. With a minimum demand level of 0.5, it was observed that the factors were proved feasible for the R1 response time, D1 standard deviation of average response time, and T1 transactions. No average size of packets (P) interval showed viable result, as seen in Figure 2:

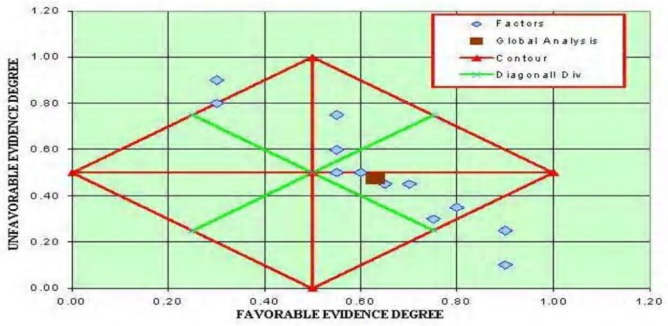

FIG. 2. Analysis of first scenario result by the Para-analyzer algorithm.

For comparison, another set of weights can be used where a higher weight is applied to each extreme position of the analyzed factor interval. The objective of this

approach is to balance the weight factor to each other while applying a slightly lower relative weight in the intermediate intervals that may generate a higher level of uncertainty, as seen in Table 5:

Table 5. Favorable and unfavorable evidences and weights of second scenario.

Factor analysis	Interval	Weight	Favorable Evidence Degree	Unfavorable Evidence Degree
Response Time	R1	2	0.9	0.1
	R2	1	0.7	0.45
	R3	2	0.55	0.75
Standard Deviation of the Average Response Time	D1	2	0.9	0.1
	D2	1	0.55	0.5
	D3	2	0.3	0.8
Average Packets Size	P1	2	0.75	0.3
	P2	1	0.65	0.45
	P3	1	0.55	0.6
	P4	2	0.3	0.9
Transactions	T1	2	0.9	0.25
	T2	1	0.8	0.35
	T3	1	0.6	0.5
	T4	2	0.3	0.8

In this second scenario, the obtained global analysis was 0.62 of favorable evidence and 0.49 of unfavorable evidence, which is slightly less than in the first scenario. With a minimum demand level of 0.5, it was observed that the factors that were viable remain the same: R1 response time, D1 standard deviation of average response time, and T1 transactions. Again, no average packets size size factor interval (P) presented viable result, as can be seen in Figure 3:

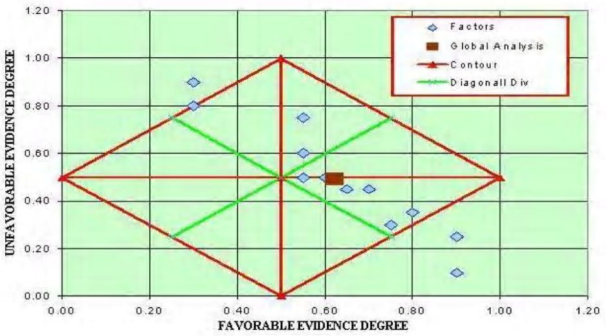

FIG. 3. Analysis of second scenario result by the Para-analyzer algorithm.

3 ANALYSIS OF THE RESULTS

From the obtained results, it can be observed that among the analyzed factors, the intervals R1, D1 and T1 gathered a common standard of viability. On the

other hand, there was no significant influence on the factor P, in any of the intervals. All the evaluated scenarios showed inconclusive results.

The interpretation of the results leads to the belief that a network with reduced response time (R1), a low standard deviation of the average response time (D1) and small number of transactions (T1) are conditions that reflect the behavior of the computer network within normal limits. However, the average size factor package does not follow the same line of reasoning, and can be proven by its own data in the log, where a significant amount of data in transit was verified with a reduced response time. Therefore, it can be concluded that the average of the data packets may not be indicative of problems in the network, only an indication of intensive use of the infrastructure.

REFERENCES

1. Lin, Y.K., Huang, C.F.: Stochastic computer network under accuracy rate constraint from QoS viewpoint. Inf. Sci. (Ny). 239, 241–252 (2013).
2. White, D., Rea, A.: A Backpropagation Neural Network for Computer Network Security. J. Comput. Sci. 2, 710–715 (2006).
3. Ben-Porat, U., Bremler-Barr, A., Levy, H.: Computer and network performance: Graduating from the "age of Innocence." Comput. Networks. 66, 68–81 (2014).
4. Kurose, J.F., Ross, K.W.: Computer Networking A Top-Down Approach Featuring the Internet. (2005).
5. Rosen, R.: Linux Kernel Networking advanced topics : Neighboring and IPsec. (2008).
6. Fossaceca, J.M., Mazzuchi, T. a., Sarkani, S.: MARK-ELM: Application of a novel Multiple Kernel Learning framework for improving the robustness of Network Intrusion Detection. Expert Syst. Appl. 42, 4062–4080 (2015).
7. Rousskov, A., Soloviev, V.: A performance study of the Squid proxy on HTTP/1.0. World Wide Web. 2, 47–67 (1999).
8. Fernandez-Prieto, J. a., Canada-Bago, J., Gadeo-Martos, M. a., Velasco, J.R.: Optimisation of control parameters for genetic algorithms to test computer networks under realistic traffic loads. Appl. Soft Comput. J. 12, 1875–1883 (2012).
9. Abe, J.M., Foundations of Annotated Logics, PhD thesis (in Portuguese) University of São Paulo, Brazil, 1992.
10. Misra, a. K., Verma, M., Sharma, A.: Capturing the interplay between malware and anti-malware in a computer network. Appl. Math. Comput. 229, 340–349 (2014).
11. Da Silva Filho, J.I., G.L.T. & J.M.A.: Uncertainty Treatment Using Paraconsistent Logic - Introducing Paraconsistent Artificial Neural Networks. (2010).
12. Da Silva Filho, J.I., G.L. Torres & J.M. Abe, Uncertainty Treatment Using Paraconsistent Logic - Introducing Paraconsistent Artificial Neural Networks, IOS Press, Holanda, Vol. 211, ISBN 978-1-60750-557-0, doi: 10.3233/978-1-60750-558-7-I, 328pp., 2010.
13. Abe, J.M., Paraconsistent logics and applications, Proceedings of 4th International Workshop on Soft Computing Applications, Arad, Romênia 1-18, ISBN 9781424479832, IEEE CFP1028D-CDR, 2010
14. Da Silva Filho, J.I. & J.M. Abe, Paraconsistent analyzer module, International Journal of Computing Anticipatory Systems, vol. 9, ISSN 1373-5411, ISBN 2-9600262-1-7, 346-352, 2001.