

A Criterion for IDS Deployment on IoT Edge Nodes

Vladimir Shakhov[1(✉)], Olga Sokolova[2], and Insoo Koo[1]

[1] University of Ulsan, Ulsan 44610, Republic of Korea
{shakhov, iskoo}@ulsan.ac.kr
[2] Institute of Computational Mathematics and Mathematical Geophysics, Novosibirsk, Russia
olga@rav.sscc.ru

Abstract. Edge computing becomes a strategic concept of IoT. The edge computing market reaches several billion USD and grows intensively. In edge computing paradigm, the data can be processed close to, or at the edge of the network. This way greatly reduces the computation and communication load of the network core. Moreover, processing data near the sources of data also provides better support for the user privacy. However, an increase in the number of data processing locations will proportionately increase the attack surface. Due to limited capacities and resources, an edge node cannot perform too many complex operations. Especially for the applications with high real-time requirements, efficiency becomes a crucial issue in secure data analytics. Therefore, it is important to get a tradeoff between security and efficiency. We focus on this problem in this paper.

Keywords: Edge computing · Internet of Things · Intrusion Detection System · Quantitative analysis · Queuing theory

1 Introduction

According to the estimation by Cisco Global Cloud Index, the data produced by IoT devices will exceed 500 Zettabytes in this year. For efficient treatment of such huge volumes of data the edge computing paradigm has been offered. In this paradigm, the data can be processed close to, or at the edge of the network. Some functions of network core is delegated to edges of the network, where the connected entities directly produce the data. These facilities can be fortified by corresponding computing platforms and system resources. Edge computing technologies greatly offload the computation and communication load of the network core. Moreover, processing data near the sources of data provides better QoS [1] for the delay sensitive services and efficient structure support for the user privacy, as well as prevent and mitigate some types of DDoS attacks.

It is expected that a ratio of enterprise-generated data, which is processed outside a conventional centralized data center or cloud, will rich 75%. The total edge computing market size is expected to grow from USD 2.8 billion in 2019 to USD 9.0 billion by 2024, at a Compound Annual Growth Rate (CAGR) of 26.5% during the forecast

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12249, pp. 546–556, 2020.
https://doi.org/10.1007/978-3-030-58799-4_40

period, estimates ResearchAndMarkets.com. According to another forecast provided by Gartner, this market will reach USD 13 billion by 2022. Financial and banking industry is one of the largest beneficiaries of edge computing worldwide. Increasing the adoption of digital and mobile banking initiatives, advanced platforms, such as blockchain and payment through mobile terminals, are fuelling the demand for modern edge computing solutions in the financial and banking industry sector. Asia-Pacific is destined to be the major market for edge computing. Businesses and governments in this region have shown more inclination toward storing and processing data locally.

However, an increase in the number of data processing locations will proportionately increase the attack surface [2]. Also, it needs to remark that we generally use edge devices with limited resources [3]. Therefore, it is not reasonable to store a large amount of data and execute a high complexity algorithm for intrusion detection. The conventional security mechanisms of cloud computing are no longer suitable for supporting security in edge computing. Taking into account security challenges, leading academia researchers and profit companies experts conclude that current situation with edge computing security is far from satisfied and essential efforts is required to overcome the existing vulnerabilities and weaknesses. Thus, edge computing security is highlighted as an important future research direction [4–6].

A lightweight and secure data analytics technique allows to increase the potential of edge computing. Due to limited capacities and resources, an edge node cannot perform too many complex operations, which could incur high latency and battery depletion. Efficiency becomes a crucial issue in secure data analytics, especially for the applications with high real-time requirements.

There are a few recent papers on the theme of Intrusion Detection System (IDS) for Edge Computing. Some authors offer a hybrid IDS mechanism, but they ignore quantitative analysis [7]. Other researchers focus on quality of detection scheme only [8]. Thus, in the present literature there is a lack of quantitative methods, which allow to form a proper holistic view on a system. This paper is intended to partially fill this gap. We describe the offered approach in general and provide a closed-for solution in an important practical case.

2 System Model

A set of IoT edge nodes serve a user-generated workload. It includes a traffic which has to be treated and retransmitted. Let us use the designations as follows.

- λ: the traffic intensity;
- μ: the intensity of request treatment;
- α: a percentage of workload of legal users, in practice, this can be estimated using an observable sample or an auxiliary model;
- B: the probability of request rejection, i.e. the blocking probability.

Here we consider a situation with two types of users. Legitimate users generate traffic with the intensity $\lambda\alpha$, and malicious users generate traffic with the intensity $\lambda(1-\alpha)$. Due to limited resources of edge nodes, a part of traffic (B) does not receive a service and rejected (see Fig. 1).

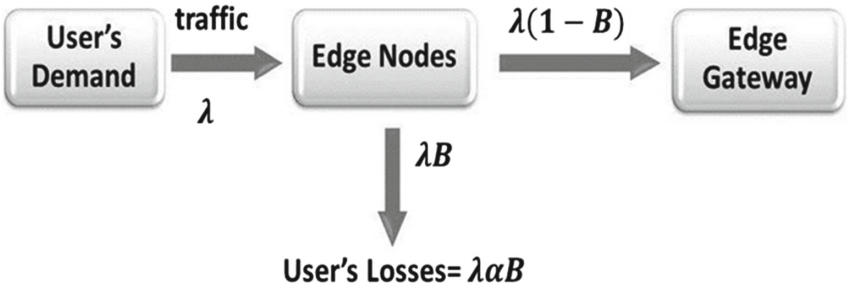


Fig. 1. IDS has not been used.

Generally, the blocking probability it is a function of λ and μ, i.e. the losses rate equals

$$\lambda B(\lambda, \mu),$$

and the served workload rate is

$$\lambda(1 - B(\lambda, \mu)).$$

Note, that not all traffic is useful. The actual losses rate of legal users is as follows:

$$l_1 = \alpha\lambda B(\lambda, \mu).$$

Next, let us consider edge nodes equipped by IDS. It is reasonable to assume that a part of malicious requests will be rejected and novel workload intensity $\widetilde{\lambda}$ will be reduced, i.e. $\widetilde{\lambda} < \lambda$. However, it does not guarantee that the system throughput becomes better. Now IoT devices have to perform additional operation for intrusion detection maintaining and malicious requests filtering. Therefore, performance of request treatment has to be reduced, i.e. the novel intensity of request treatment becomes $\widetilde{\mu}$, and $\widetilde{\mu} < \mu$.

If the security system is designed to counteract a limited set of known attacks, then signature based IDS can be used. In this case IDS utilizes a set of rules (signatures) that can be used to detect an attacks pattern. This approach provides a high level of accuracy for well-known intrusions. A signature based IDS is characterized by low computational cost, i.e. $\widetilde{\mu} \approx \mu$. The same effect can be reached by the use of small number of secret bits for requests verification. However, this situation is not typical for IoT environment. Intruders constantly change tactics and create new destructive tools. Signature-based detection does not detect slightly modified attacks, much less it does not detect unknown attacks. Hence, advanced intrusion detection methods are required.

Also, the situation when $\widetilde{\mu} \gg \mu$ is not typical for IoT keeping in mind the edge devices level [9]. Low resources make ineffective heavy computation algorithms like deep learning. So, it is reasonable to assume that performance of request treatment has not been increased drastically.

Moreover, some legitimate requests are mistakenly recognized as illegal and filtered by IDS (Fig. 2).

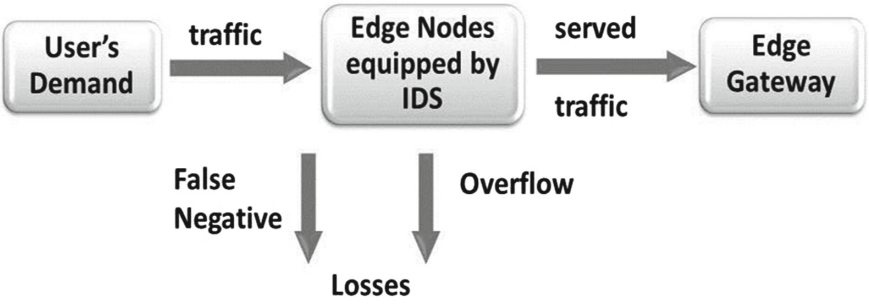


Fig. 2. Nodes are quipped by IDS.

Let us investigate the situation, where IDS deployment makes a sense.

3 Analysis

3.1 IDS Applications

For our purposes it is enough to consider IDS parameters as follows:

- p_I is a false positive, i.e. the probability of the event, when a legitimate request is rejected by IDS;
- p_{II} is a false negative, i.e. the probability of the event, when an illegal request is accepted.

Therefore, IDS rightly rejects

$$\lambda(1-\alpha)(1-p_{II})$$

spoofed request per time unit. And the forced losses of legal workload are as follows:

$$\lambda\alpha p_I.$$

Hence, edge nodes have to treat an offered load of intensity:

$$\widetilde{\lambda} = \lambda\alpha(1-p_I) + \lambda(1-\alpha)p_{II}$$

Let us remark, that the ratio of legitimate requests has been changed. Now, this one is

$$\widetilde{\alpha} = \frac{\lambda\alpha(1-p_I)}{\widetilde{\lambda}} = \frac{\alpha(1-p_I)}{\alpha(1-p_I)+(1-\alpha)p_{II}}.$$

In the case of IDS application, the actual losses rate of legal users becomes as follows:

$$l_2 = \lambda\alpha p_I + \widetilde{\alpha}\widetilde{\lambda}B\left(\widetilde{\lambda}, \widetilde{\mu}\right).$$

The IDS (with given quality parameters p_I, p_{II}) is justified if and only if

$$l_2 < l_1.$$

Hence, the novel intensity of request treatment has to satisfy the inequality as follows:

$$B\left(\widetilde{\lambda}, \widetilde{\mu}\right) < \frac{\alpha\lambda}{\widetilde{\alpha}\widetilde{\lambda}}\left(B(\lambda, \mu) - p_I\right) \tag{1}$$

In view of the above considerations, inequality (1) can be given an alternative formulation as follows:

$$B\left(\widetilde{\lambda}, \widetilde{\mu}\right) < \frac{B(\lambda, \mu) - p_I}{1 - p_I}.$$

Using an appropriate queuing model we can calculate the blocking probability and its inverse.

3.2 Erlang-B Function

Let us specify the model. Taking into account requirements of delay sensitive services, it is reasonable to use M/M/n/n queuing system to model functioning of cluster containing n edge nodes. Thus, our assumptions are as follows:

- Incoming Poisson flow (with intensity $\widetilde{\lambda}$ or λ)
- Exponential service time (with intensity $\widetilde{\mu}$ or μ).
- No buffer (waiting room)

In this case the blocking probability is described by Erlang-B formula (see, for example [10]),

$$B(\rho, n) = \frac{\frac{\rho^n}{n!}}{\sum_{i=0}^{n} \frac{\rho^i}{i!}},$$

where

$$\rho = \frac{\lambda}{\mu}.$$

The inequality (1) can be numerically solved. Note, the assumption of exponential CDF for service time is not necessary. The formula is true for M/G/n/n queuing system as well.

Let us consider the case, when

$$\rho \gg n.$$

It generally takes place under attack. We use the following theorem [11].

Theorem.

$$\forall \varepsilon > 0, \text{ if } \rho \geq n + \frac{1}{\varepsilon},$$

then

$$1 - \frac{n}{\rho} < B(n, \rho) < 1 - \frac{n}{\rho} + \varepsilon.$$

Corollary. If ε is small enough, then we get an approximation for Erlang-B function:

$$B(n, \rho) \approx 1 - \frac{n}{\rho},$$

the inverse functions approximations can be easily calculated as well

$$\rho = \frac{n}{1 - B},$$

$$n = \rho(1 - B).$$

Thus, M/M/n/n system under heavy load provides the outgoing rate (served requests) as follows:

$$\lambda(1 - B(n, \rho)) = \lambda \frac{n}{\rho} = n\mu,$$

and the losses rate is as follows:

$$\lambda B(n, \rho) = \lambda - n\mu.$$

3.3 Service Differentiation

Let us consider the problem of service differentiation. It can be a problem of security differentiation for different classes of customers as well as a traffic management problem, cluster head selection process etc. In the case of jamming attacks, this technique can be used to assign non-attacked channels in order to support the survivability of the most important applications.

The problem statement is as follows

$$N \to min$$

$$N = \sum_{j=1}^{C} n_j$$

$$B\left(n_j, \rho_j\right) \leq b_j, j \in \{1, 2, .., C\},$$

where

- N is a total number of computational resources (channels, servers, service centres, IDS agents etc.)
- C is a number of user classes.
- n_j is a number of resources assigned to the class j
- b_j is QoS required by class j (i.e. the losses rate).

In the case of limited resources (the most critical case) the approximation above helps to solve the problem. The solution is as follows

$$n_j = \rho_j(1 - b_j), j \in \{1, 2, .., C\},$$

$$N_{opt} = \sum_{j=1}^{C} \rho_j(1 - b_j).$$

4 Criterion

Now we can get a closed-form solution for the inequality (1) in the case of heavy workload. It is as follows.

Proposition. IDS is justified if the following inequality is true:

$$\widetilde{\mu}(1 - p_I) > \mu(\alpha(1 - p_I) + (1 - \alpha)p_{II}) \tag{2}$$

This inequality can be used for estimation and selection intrusion detection algorithms. For convenience, the inequality (2) can be rewritten as a ratio of requests processing intensities:

$$\frac{\widetilde{\mu}}{\mu} > \alpha + \frac{p_{II}}{1 - p_I}(1 - \alpha).$$

We can often (but not always) expect that a way to improve the false positive parameter entails the consequences of the proportional degradation of the false negative parameter, and vice versa. This is a specific of IDS design. However, if the IDS quality is good enough, both p_I and p_{II} are small enough.

Let us define the following ratio:

$$\frac{p_{II}}{1-p_I},$$

as the IDS throughput. Generally, packets are processed individually by IDS, hence this value does not depend on the legal users' packets proportion.

Let us remark, if the efficiency of used intrusion detection algorithms is very high, i.e.

$$p_I \to 0,$$

$$p_{II} \to 0,$$

then the criterion for the appropriateness of IDS takes a simple form:

$$\alpha < \frac{\widetilde{\mu}}{\mu} < 1.$$

The decision to deploy IDS (or provide requirements for IDS) can be based on a profitability analysis. Therefore, a criterion can take a set of various forms. For example, the criterion can be as follows: IDS should improve the loss rate by k times, i.e.

$$\frac{l_2}{l_1} > k,$$

where k is a desired constant. An alternative statement can be as follows: the IDS profitability has to be higher than the desired threshold k, i.e.

$$l_2 > k.$$

Next, it needs to maximize the system profitability under limited cost or energy consumption, i.e.

$$\max(l_1, l_2)$$

$$\widetilde{\mu} \leq k,$$

and so on. The approximation above allows to obtain a closed form solution in these cases as well.

5 Performance Evaluation

In our consideration we can assume that the IDS throughput value varies in the range [0 … 1]. The following equation can be useful to determinate a tradeoff between computational overhead and intrusion detection efficiency

$$\frac{\widetilde{\mu}}{\mu} = \alpha + (1-\alpha)\frac{p_{II}}{1-p_I}$$

Figure 3 presents the increase in IDS overhead change according to IDS throughput efficiency for $\alpha \in \{0.1; 0.5; 0.9\}$.

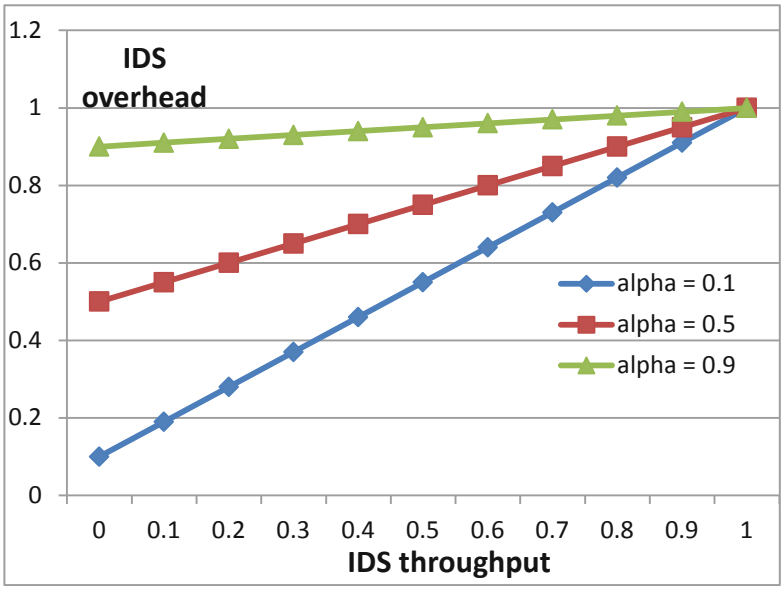


Fig. 3. The increase in IDS overhead change according to IDS throughput.

As we can see, if the portion of legitimate requests is about 10%, and IDS leads to a decrease in node performance by 50%, then the throughput of IDS can confidently be in the range from 0.2 to 0.3. This is a very mediocre IDS. And next example, if the portion of legitimate requests is about 90%, then no reasons to use even an ideal IDS with no any mistakes (zero false positive and false negative) and only 15% degradation of node throughput.

Let us consider the value as follows:

$$\frac{\frac{\widetilde{\mu}}{\mu} - \alpha}{\alpha} * 100\%.$$

Assume, the false positive and false negative values are small enough. Here, without loss of generality we can take $p_I = p_{II} \in \{1\%, 5\%, 10\%\}$.

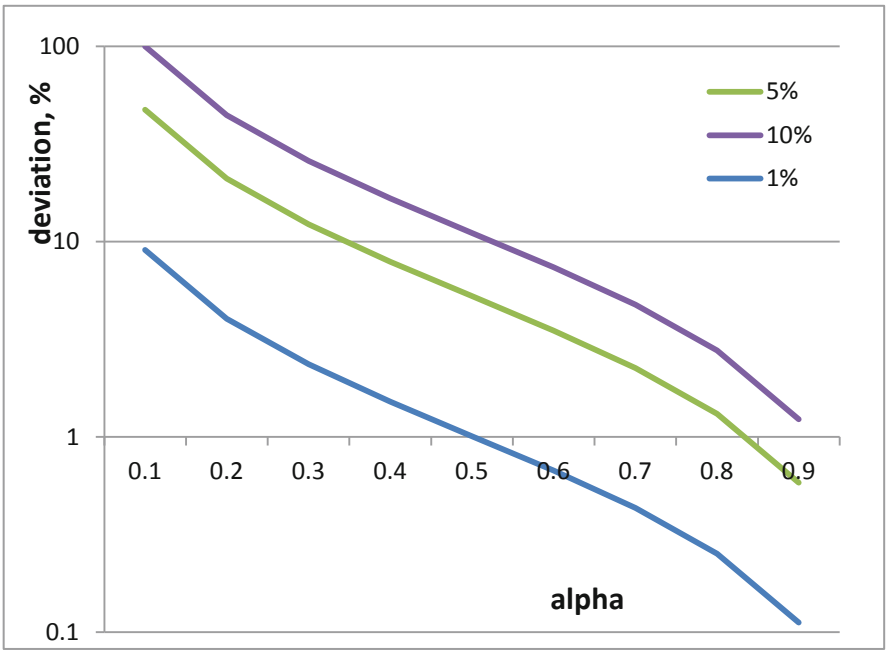


Fig. 4. The suitability of α as the node performance degradation threshold depends on the quality of the used intrusion detection algorithm.

If the quality of the used intrusion detection algorithm is very high (about percent or less), then α can be taken as a threshold for reducing node performance due to IDS operation. This is not true if the values p_I and p_{II} exceed 5%, although this is still a good enough intrusion detection algorithm. Figure 4 illustrates this point.

6 Conclusion

In this paper we offer a criterion for IDS deployment on IoT edge nodes. The offered results are based on queuing theory. In particular we use M/M/n/n (M/G/n/n) systems. In general the provided approach can be applied to any kind of IDS. However, the detailed results are provided for the low-resources IoT devices (Edge nodes). Using the Erlang losses function approximation we received a quantitative condition when IDS deployment makes a sense. The offered approach can be mostly applied for situations of flooding-type intrusions. Note, that the result can be used in other application domains such as enterprises management, hospital operation and others.

Acknowledgment. This work was supported by the National Research Foundation of Korea (NRF) grant through the Korean Government (MSIT) under Grant NRF-2020R1I1A1A0 1065692. The reported study was partially supported by RFBR according to the research project 19-01-00562.

References

1. Zhang, J., Chen, B., Zhao, Y., Cheng, X., Hu, F.: Data security and privacy-preserving in edge computing paradigm: survey and open issues. IEEE Access **6**, 18209–18237 (2018)
2. Xiao, Y.: Edge computing security: state of the art and challenges. Proc. IEEE **107**(8), 1608–1631 (2019)
3. Liu, F., Tang, G., Li, Y., Cai, Z., Zhang, X., Zhou, Y.: A survey on edge computing systems and tools. Proc. IEEE **107**(8), 1537–1562 (2019)
4. Taleb, T., Samdanis, K., Mada, B., Flinck, H., Dutta, S., Sabella, D.: On multi-access edge computing: a survey of the emerging 5G network edge cloud architecture and orchestration. IEEE Commun. Surv. Tut. **19**(3), 1657–1681 (2017)
5. Mao, Y., You, C., Zhang, J., Huang, K., Letaief, K.: A survey on mobile edge computing: the communication perspective. IEEE Commun. Surv. Tut. **19**(4), 2322–2358 (2017)
6. Porambage, P., Okwuibe, J., Liyanage, M., Ylianttila, M., Taleb, T.: Survey on multi-access edge computing for Internet of Things realization. IEEE Commun. Surv. Tut. **20**(4), 2961–2991 (2018)
7. Yao, H., et al.: Hybrid intrusion detection system for edge-based IIoT relying on machine-learning-aided detection. IEEE Networks **33**(5), 75–81 (2019)
8. Lin, F., et al.: Fair resource allocation in an intrusion-detection system for edge computing: ensuring the security of Internet of Things devices. IEEE Consum. Electron. Mag. **7**(6), 45–50 (2018)
9. Jan, S., Ahmed, S., Shakhov, V., Koo, I.: Toward a lightweight intrusion detection system for the Internet of Things. IEEE Access **7**, 42450–42471 (2019)
10. Medhi, J.: The evergreen Erlang loss function. OPSEARCH **43**(3), 309–319 (2006). https://doi.org/10.1007/BF03398780
11. Shakhov, V.V.: An efficient method for proportional differentiated admission control implementation. In: Vinel, A., Bellalta, B., Sacchi, C., Lyakhov, A., Telek, M., Oliver, M. (eds.) MACOM 2010. LNCS, vol. 6235, pp. 91–97. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15428-7_10