Annals of Telecommunications
https://doi.org/10.1007/s12243-021-00894-4

A balanced prior knowledge model based on Beta function for evaluating DIDS performance

A modeling update

Renato S. Silva[1] · **Luís F. M. de Moraes**[1]

Received: 21 April 2021 / Accepted: 21 October 2021
© Institut Mines-Télécom and Springer Nature Switzerland AG 2021

Abstract
A federation-based DIDS is a security platform composed of autonomous IDS able to learn with their data and cooperate with each other to improve the overall detection performance. However, evaluating the detection performance of a DIDS, specially considering its heterogeneous environment and the wide range of threats that emerge every single day, is not trivial. Although the Bayesian inference approach presents itself as a compatible option to model this kind of systems, lacking a sufficiently large and diverse dataset is a relevant issue for building blocks of prior knowledge. Our approach relies on the "learn-from-data" insight of the Beta function to propose a modeling framework aiming to assess the overall detection performance of DIDS systems, regardless of dataset rounds. Comparing our results to the numbers obtained either from testbeds or simulation, the proposed model presents a fair approximation.

Keywords DIDS · Probability of detection · Beta distribution · Functional performance evaluation

1 Introduction

An intrusion detection system (IDS) attempts to identify unauthorized use, misuse, or abuse of computer and network systems. As intrusions become even more sophisticated, dealing with them requires increasingly complex and high-performance systems [1], able to consistently cover an ever-growing wide range of threats. The assessment of the functional performance of an intrusion detection system (IDS) depends on metrics especially designed to measure its detection amplitude and its auditing capacity to distinguish intrusive from non-intrusive activities [2]. These metrics are based on numbers that are systematically organized in a framework named confusion matrix, containing four possible results: false-positive (#FP), false-negative (#FN), true-positive (#TP), and true-negative (#TN). However, building a confusion matrix depends on empirical tests of training the IDS in question using a dataset containing both normal and malicious data. Besides choosing a sufficiently large and diversified dataset, its construction process should match the network scenario in which this IDS is supposed to operate. Even using a simulation model, the need for building a confusion matrix from a dataset training remains once the performance metrics keep coming from its numbers. Table 1 shows the confusion matrix framework.

According to [3], there are thousands of different public datasets on the Internet to download, such as DARPA98, KDD99, ISC2012, and ADFA13. Although there are many options, generally, datasets are built using particular proceedings that consider some specific objectives they were meant for [4]. In other words, there is not a complete dataset that comprises all the known types of intrusions, without taking into account those that are still unknown. Besides, some of these datasets may lack traffic diversity and volume, while others contain anonymized packet information, which cannot fairly reflect the current trends [5]. It means, even testing the same IDS, each dataset gives rise to different confusion matrix, and consequently to distinct evaluation outcomes regarding its performance [6]. This problem becomes even more challenging in the case of distributed intrusion detection systems (DIDS) composed of several independent IDS members geographically spread across the worldwide Internet.

A distributed intrusion detection system (DIDS) can be considered as an evolution of typical IDS aiming to improve functional performance and overcome problems related

✉ Renato S. Silva
renato@ravel.ufrj.br

[1] Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Table 1 Confusion matrix disposition

IDS	Intrusion	Normal
Detection	#TP	#FP
No detection	#FN	#TN

to its monolithic perimeter-based architecture. According to [7], a DIDS consists of multiple individual intrusion detection systems (IDS) spread on a network, where all of which communicate with each other or with a central server to facilitate advanced network monitoring, incident analysis, and instant attack data. The grounding insight behind the DIDS approach relies on expanding the underlying sensory surface by combining information from multiple sources to increase the detection range (eventually including zero-day attacks[1]) and reduce the false-positive alarms rate. However, evaluating the functional performance of a DIDS composed of autonomous agents using empirical numbers obtained from rounds of data sets is a hard-boiled task and may not reproduce such a heterogeneous and dynamic environment. In this case, building an analytic framework to model the DIDS behavior can be considered as an alternative option to approximate the practical scenario. Central to the Bayesian philosophy lies in describing all the unknown data probabilistically, even before observing them, as in the prior probabilities. Nonetheless, specifying a prior that fairly reproduces the previous knowledge regarding any system remains a challenge due to the subjectivity it sometimes involves [8].

In this paper, we consider the DIDS platform described in our anterior work [9] to propose a novel analytic model blending Bayesian inference with Beta distribution aiming to evaluate the functional performance of DIDS platforms. This paper also contributes with a new analytical framework reproducing a practical DIDS scenario, in which autonomous detection agents geographically distributed cooperate with each other as a detection federation to alarm potential security threats. Instead of using static prior knowledge obtained empirically, which tends to be out of date over time, the proposed model reduces the dependence on a dataset-based confusion matrix by dynamically balancing the Bayesian likelihood with the current intrusion evidence. In other words, besides updating the likelihood parameter by combining the belief degree of the arriving advertisements, the prior knowledge about the detection probability of the DIDS platform fits the current detection scenario. To validate the modeling approach, we present a comparative analysis involving the two models considered in this paper with the results obtained from a testbed. The performance results obtained from the metrics modeled according to the proposed approach show a more coherent behavior considering the trade-off involving the sensor network size and the number of intrusion evidence received at the combination point.

The rest of the paper is organized as follows. In Section 2, we present the main works focused on evaluating the detection performance of distributed intrusion detection systems. We also emphasize the contribution of this paper by comparing the approaches. Section 3 reviews the architecture of the DIDS platform considered in this paper. In Section 4, we present our modeling approach to evaluate the detection performance of a consensus-based DIDS. Section 5 presents a comparative analysis between the analytic models considered in this paper. In Section 6, we close the article with an objective analysis correlating the results obtained from the models with the paper contribution.

2 Related works

It is not new the idea of distributing multiple IDS and making them collaborate with each other to improve functional performance and overcome critical problems related to the monolithic perimeter-based architecture of typical IDSes. Indeed, such a visionary approach keeps being the central matter of several works, such as [10–14], a long time ago. Some of these works, and many others, still ground more recent proposals motivated by the drastic increase of the number and the complexity of new cyberattacks. It is the case in [15], where the authors propose a distributed architecture composed by lightweight IDSes that hierarchically collaborate with each other to overcome the single point of failure. The results obtained from dataset training using different compositions show notable improvements when compared to the non-collaborative model.

More recently, the work presented in [16] extends the concept of Intrusion Detection Networks (IDN) as an overlay network that enables IDSes to exchange intrusion information and knowledge in order to improve the overall performance. The IDN architecture proposed by Fung et al. is based on a collaboration framework that includes both trust and acquaintance managements, aiming to prevent dishonest or incompetent nodes and insider attacks. The functional performance of the collaborative decision system is modeled by associating a acquaintances cost for each of the four possible outcomes of the confusion matrix, where the Beta distribution combined with Bayesian inference is used to model the positive rate of each IDS member. The authors compare the results obtained from the proposed model with

[1] A named zero-day attack occurs when an attacker exploits a vulnerability before security teams can find a fix.

other heuristic approaches (simple and weighted average aggregation) to evaluate the overall detection performance of the IDN approach. Finally, the robustness and scalability of the system are also assessed by analyzing the impact of the malicious insiders in the system performance, and the relationship between the number of acquaintances versus the number the nodes of the network, respectively.

Closely related to the collaborative essence of distributed intrusion detection systems is the concept of federated learning. It permits multiple actors to build a common machine learning model without sharing user data, thus addressing critical issues such as data privacy, data security, data access rights access to heterogeneous data. Intrusion detection approaches like in [17] and [18] used federated learning algorithms to solve specific problems related to privacy and resource-constrained environments to propose distributed intrusion detection systems with high functional performance.

The Bayesian philosophy refers to combining the evidence contained in a signal with the prior knowledge of the probability distribution of the process. The work presented in [19] proposes a method based on Bayesian estimation to sequentially update the probability of detection for tracking. The proposed approach forms a feedback loop where the tracker output is used to estimate the prior probability θ after each scan. In their analytic model, the authors assume $\Pr(\theta)$ as a Beta distribution with parameters (α, β). The up to date value for θ is obtained by computing the mean conditional probability $E_{\Pr(\Theta|X_k)}[\Theta]$, being X_k the data obtained in the k^{th} scan. The simulation results indicated that using high informative prior decreases the detection performance.

Due to the variety of shapes it can take, Beta distribution is commonly used to reproduce semi-informative prior probabilities in case of Bayesian anomaly detection. The analytic framework proposed in [20] relies on a univariate time series algorithm to detect the occurrence of an anomalous disease outbreak. In their proposal, the authors use a dataset as a baseline to evaluate the Beta parameters α_0 and β_0, assuming θ_{OB} as a random variable with uniform density distribution $f(\theta_{OB})$ representing the fraction of people that visited the emergency department with a chief complaint. To evaluate the functional performance of the proposed system, the authors compared the results obtained from a real 12-week dataset from a set of emergency departments in Allegheny Country, Pennsylvania in 2001 and 2002, with the control chart method. The activity monitoring operating characteristic (AMOC) curves show that the proposed Bayesian univariate (BU) approach has relatively better functional performance when using baseline periods in which the window size $w = [2 - 10]$ weeks.

In our previous work [9], we propose a distributed intrusion detection platform able to detect any kind of attack/intrusion that uses the Internet connectivity to reach their targets, such as any variation of denial-of-services (DoS) attacks, worms and network scanning. The proposed architecture leverages the scope of the BGP network to interconnect IDS members operating at the autonomous systems of the Internet (AS), enabling them to cooperate with each other by sending network alarms as a federation. A federated IDS member traversed by a supposed malicious flow detects it as a potential intrusion and advertises a BGP FlowSpec updating message containing the traffic information (features) [21]. Each intermediate AS that receives the update message relays it immediately, causing it to travel through the BGP network until it reaches the destination AS. At the destination AS, all messages arriving from multiple sources are correlated and combined to support further protection measures, depending on their resulting degree of belief. The overall functional performance of the DIDS platform depends on the individual performance of each federated IDS and is evaluated based on detection metrics modeled by combining the Dempster-Shafer Theory of Evidence with Bayesian inference. In those models, the positive-prediction value (PPV) related to each originating IDS assigns a belief mass to its BGP advertisement. At the target AS, the correlated belief masses are mathematically combined to yield the likelihood parameter in the Bayesian equation. However, although the detection numbers from the sensory surface continuously update the likelihood parameter, the prior knowledge regarding the empirical detection probability of the DIDS keeps flat. Besides the unrealistic scenario of training the whole DIDS platform using sufficiently large and diversified datasets, this empirical method does not consider the autonomy of the AS administrator to use state-of-art IDS members, for example, the ones that progressively improve their own functional performance by learning with the data, as proposed in [22, 23].

The modeling approach proposed in this paper leverages the multi-shaping nature of the Beta distribution to balance the likelihood parameter of the metric models concerning the current intrusion evidence. This approach fits better the practical scenario, in which the detection probability of each individual IDS that composes the DIDS federation changes continuously instead of keeping flat over time. Therefore, instead of modeling adjacent aspects related to the overall performance, as proposed in [16], or relying just on empirical results from historical data, as suggested in [9, 20], we present a modeling approach that uses the intrusion evidence as input parameters to compose the prior probability. Our main contribution in this paper relies on proposing a simple analytic framework based on the Beta distribution as a prior probability to evaluate the overall functional performance of a consensus-based DIDS that reduces the dependence over dataset training.

3 Architecture review

The same native ubiquitous connectivity offered by the Internet that enables cyberattackers to launch a malicious attempt against any target from anywhere also paves the way for the collaboration-based distributed intrusion detection system proposed in [9]. The system proposed in our previous paper works like an intrusion detection network, where IDS members placed in every autonomous system (AS) collaborate with each other by sharing specific information about a potential malicious flow that traverses them towards a target AS. Each intrusion advertisement from an IDS member traversed by the malicious flow travels throughout the BGP network from its source AS and arrives at the target AS, where it is correlated and combined with other correlated messages to support a protection measure. The longer the AS-path of a malicious flow, the bigger the chances for at least one IDS member in the AS-path to detect it as an intrusion. If one considers a coordinated attack from multiple sources spread across the worldwide Internet, the detection probability exponentially increases.

The proposed architecture assumes each AS has at least one federated IDS, which is autonomously controlled by the AS administrator and can be of any type (e.g., an anomaly-based or signature-based like Snort) with different functional performance. The heterogeneity and autonomy of IDS members placed in their respective ASes increase the chances of detecting a zero-day attack, since it may have been identified as an attack earlier in a different place. Besides that, the federated IDSes are also assumed independent of each other. Therefore, any individual IDS that fails does not significantly compromise the overall DIDS performance. Furthermore, the multiple detection messages from the DIDS federation correlated and combined at the target AS corroborate the intrusion hypothesis, reducing the false-positive rate and increasing the true-positive rate of the platform as a whole.

The lightweight self-organized BGP network interconnects all the autonomous systems (AS), including their respective distributed IDSes, creating a global intrusion detection federation. The twelve flow attributes of the standardized BGP Flowspec framework normalize the communication among the IDS members, without restricting the flow characterization, which facilitates the correlation process at the destination AS (Fig. 1).

To evaluate the functional performance of the DIDS platform described in [9], we proposed an analytic model based on combining the belief masses of the N_D advertisements into consolidated information to support further protection measures. However, using a flat positive rate empirically obtained from dataset rounds to model the prior knowledge about the system's capacity to detect an intrusion disregards the trade-off between the federation size and the positive precision of IDS.

4 Modeling framework

As mentioned in Section 2, the modeling approach presented in our previous work [9] proposes to use the Dempster-Shafer framework for combining the belief masses of the correlated detection messages from the distributed IDSes about a potential intrusion in progress. The combined result enters the Bayesian statistical models as the likelihood parameter to assess performance metrics. The belief mass (or basic probability assignment) concerning each message that arrives to be combined at the destination AS is parameterized using the positive-prediction value (PPV) of the source IDSes, which is related to their ability to alarm in case of a real intrusion in progress. This modeling approach is capable of capturing both the impact of the number of messages agreeing that there is an intrusion and the average precision of each IDS member. However, the prior knowledge that measures the DIDS ability to detect an intrusion, which can be true or false, depends only on the individual positive rate (PR_{av}) of each federated IDS, which is empirically evaluated by averaging two different datasets trained in the same standardized IDS, namely Snort.

$$PR_{av} = \frac{1}{JN_F} \sum_{i=1}^{N_F} \sum_{j=1}^{J} \frac{\#TP_{ij} + \#FP_{ij}}{\#TP_{ij} + \#FP_{ij} + \#TN_{ij} + \#FN_{ij}} \tag{1}$$

In Eq. 1, N_F refers to the total number of IDS members of the detection federation and J is the number of datasets used to train our chosen IDS baseline (Snort).

Although the performance metrics in [9] present a coherent behavior considering the federation size and the number of intrusion evidence, using just two datasets is not consistent enough to evaluate the detection performance of the whole DIDS platform. Also, the flat prior knowledge from the positive rate does not consider the possibility of changes in the IDS performance over time. By the way, most modern IDSes use artificial intelligence and machine learning techniques to improve their detection algorithms using their own traffic.

This section proposes a different modeling approach aiming to highlight the numerical impact of the detection evidence and reduce the model's dependence on the dataset rounds of training. Although from a classical perspective it is unacceptable to place probability distributions on parameters (frequentist point of view), we propose a Bayesian approach to estimate the prior, in which the model parameter is represented with a probability distribution, whereas the data already observed are considered as fixed numbers [24]. However, unlike the maximum likelihood (ML) and maximum a posterior (MAP) approaches, whose objectives are to maximize the likelihood or the posterior with the data [25], we propose a balanced semi-informative prior as a random

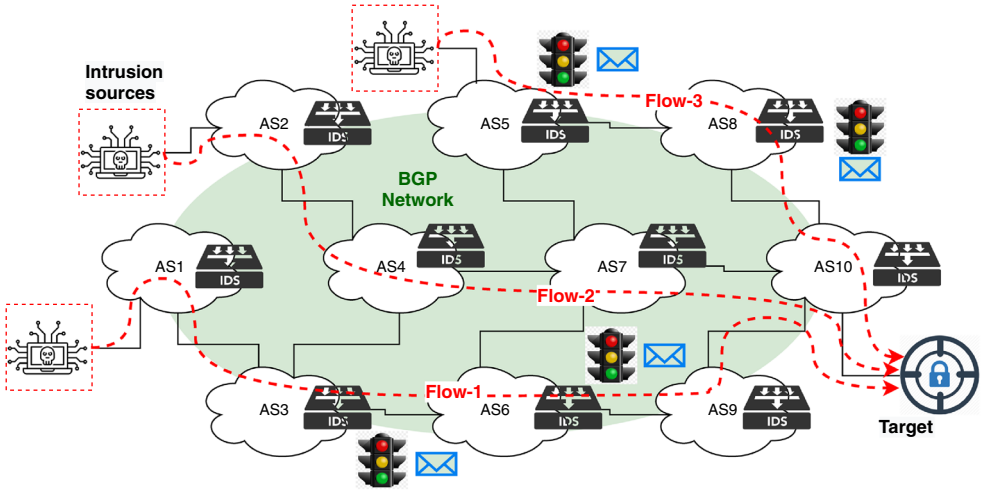


Fig. 1 Distributed detection architecture with 10 ASes, each of them containing at least one federated IDS member. A coordinated attack is launched from ASs 1, 2, and 5 towards a target placed at AS 10. Four IDSes along the AS-paths detect the malicious flows and advertise a FlowSpec update to the BGP network

variable, blending historical and current data to allow performance model to learn with time. Figure 2 gives a graphical view for a better understanding of the proposed approach.

In Fig. 2, the statement about the occurrence of intrusions can be modeled by a Bernoulli random variable I, where $\Pr(I = 1) = 1 - \Pr(I = 0)$ is the probability of there is a real intrusion ongoing in the network at a time instant t. Similarly, let U_i an individual Bernoulli random variable representing the detection statement of an IDS$_i$, as part of the DIDS federation, where $\Pr(U_i = 1) = 1 - \Pr(U_i = 0)$ is the probability of an IDS$_i$ to detect an intrusion.

TPR$_{ij}$ (True-Positive Rate) is a detection performance metric that measures the IDS sensitivity to detect a real intrusion. It can be evaluated by rating the number of true-positive alarms issued by IDS_i (#TP) over the number labeled intrusion flows in the dataset j. In other words, considering the total amount of malicious flows present in a dataset j ($\#TP_{ij} + \#FN_{ij}$), TPR$_{ij}$ measures the capacity of an IDS$_i$ to detect these intrusive flows.

$$TPR_{ij} = \frac{\#TP_{ij}}{\#TP_{ij} + \#FN_{ij}} \tag{2}$$

In Eq. 2, both $\#TP_{ij}$ and $\#FN_{ij}$ can be evaluated based on the confusion matrix CM_{ij}. While $\#TP_{ij}$ is the number intrusions correctly alarmed by the IDS_i, $\#FN_{ij}$ is the number of real intrusions not detected by IDS$_i$, trained using the dataset j.

Assuming j a sufficiently diversified and voluminous dataset, and identical independent IDSes in the federation, it

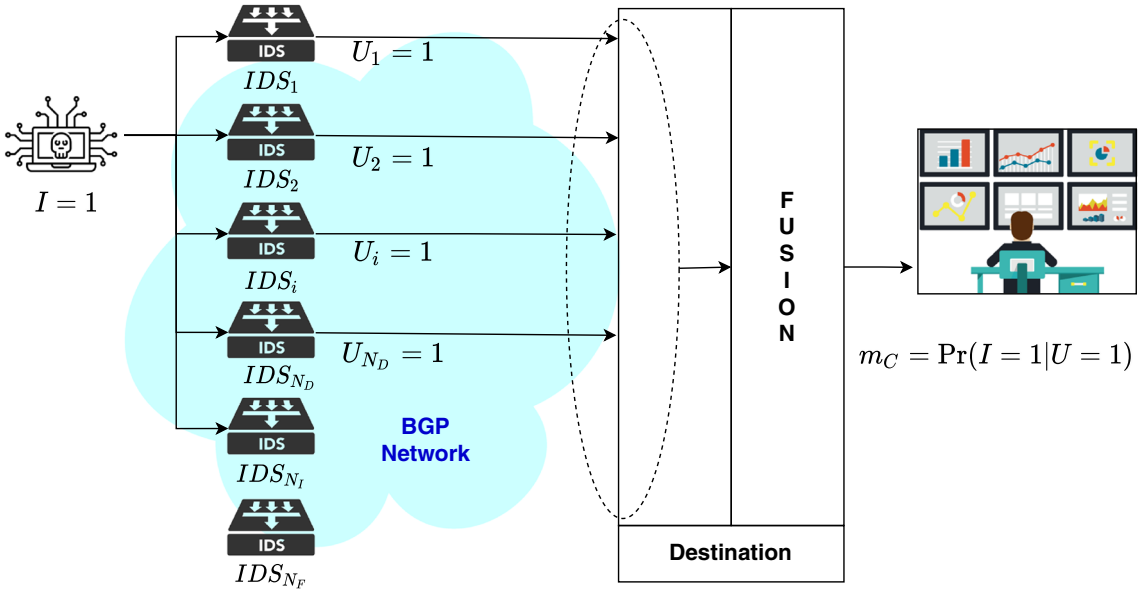


Fig. 2 Graphical model representing the intrusion hypothesis ($I = 1$) the N_D detection hypothesis ($\{U_i\}_{i=1}^{N_D} = 1$)

is possible to aproximate both PR_i and TPR_i to their statistical model.

$$PR_i = \Pr(U_i = 1) \tag{3}$$

$$TPR_i = \Pr(U_i = 1 | I = 1) \tag{4}$$

Using the Bayes framework, we can rewrite TPR_i from Eq. 4 into Eq. 5.

$$TPR_i = \Pr(U_i = 1 | I = 1) = \frac{\Pr(I = 1 | U_i = 1) \times \Pr(U_i = 1)}{\Pr(I = 1)} \tag{5}$$

Equation 5 denotes the detection metric named true-positive rate of an averaged standardized IDS_i in the detection federation, and can be interpreted as the posterior probability, according to Eq. 6.

$$Posterior = \frac{Likelihood \times Prior}{Evidence} \tag{6}$$

Comparing Eq. 5 to Eq. 6, $\Pr(U_i = 1) = p$ plays a role of the prior probability, which represents the previous belief about the positive detection capacity of the IDS_i, before taking into account any data evidence. In Bayesian inference, it is crucial to the prior probability to fit the likelihood, according to the system behavior. Otherwise, the posterior probability can turn out meaningless values. That is, while a very informative prior probability can pull the posterior outcome to meaningless area, a non-informative prior increases the variance of the posterior evaluation [26].

The modeling approach proposed in this paper focus on forming a semi-informative prior as a function of the detection scenario. In other words, instead of using a flat prior probability $p = PR_{av}$, which depends on the unthinkable task of constantly training each federated IDS using the same dataset, let p a random variable $\Theta \in [0, 1]$ with $F_\Theta(\theta) \in [0, 1]$.

Assuming $X \leq K$ be random variables representing the number of IDSes that detected an intrusion and the number of IDSes traversed by the malicious flow, respectively, the likelihood $\Pr(X = N_D | K = N_I, \Theta = \theta)$ can be written in Eq. 7.

$$\Pr(N_D | N_I, \theta) = \binom{N_I}{N_D} \theta^{N_D} (1 - \theta)^{N_I - N_D} \tag{7}$$

As one can see, Eq. 7 has a binomial shape. Therefore, choosing a Beta distribution as the prior probability only changes the distribution parameters (hyper-parameters) due to its binomial conjugate property. However, defining how informative the prior should be depends on adjusting its shape parameters α and β accordingly. In our case, we focus on the number of correlated messages to be combined N_D and on the proportional size of the federation N_I, which represent the real detection scenario at the attack target. The Beta probability density function $f_\Theta(\theta)$ follows next.

$$f_\Theta(\theta) = \frac{dF_\Theta(\theta)}{d\theta} = \frac{1}{B(\alpha, \beta)} \theta^{\alpha - 1} (1 - \theta)^{\beta - 1} \tag{8}$$

Where, $B(\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}$

According to Bayes theorem, the posterior $\Pr(\theta | N_D, N_I)$ is

$$\Pr(\theta | N_D, N_I) = \frac{\Pr(N_D, N_I | \theta) f_\Theta(\theta)}{\Pr(N_D, N_I)} \tag{9}$$

Joining Eqs. 7 and 9, we have the following expression for the posterior probability.

$$\Pr(\theta | N_D, N_I) = \frac{\binom{N_I}{N_D} \theta^{N_D} (1 - \theta)^{N_I - N_D} \theta^{\alpha - 1} (1 - \theta)^{\beta - 1}}{B(\alpha, \beta) \int_0^1 \Pr(N_D, N_I | \theta) f_\Theta(\theta) d\theta} \tag{10}$$

Developing the integral in the denominator of Eq. 10, we have

$$\begin{aligned} \int_0^1 \Pr(N_D, N_I | \theta) f_\Theta(\theta) d\theta &= \int_0^1 \frac{1}{B(\alpha, \beta)} \binom{N_I}{N_D} \theta^{N_D + \alpha - 1} (1 - \theta)^{N_I - N_D + \beta - 1} d\theta \\ &= \frac{1}{B(\alpha, \beta)} \binom{N_I}{N_D} B(N_D + \alpha, N_I - N_D + \beta) \end{aligned}$$

Coming back with this last result in Eq. 10

$$\Pr(\theta | N_D, N_I) = \frac{\theta^{N_D + \alpha - 1} (1 - \theta)^{N_I - N_D + \beta - 1}}{B(N_D + \alpha, N_I - N_D + \beta)} \tag{11}$$

Taking the marginal concerning N_D, we have the posterior

$$\begin{aligned} \Pr(\theta | N_I) &= \sum_{x=0}^{N_I} \Pr(\theta | N_D = x, N_I) \Pr(N_D = x | N_I) \\ &= \frac{\theta^{x + \alpha - 1} (1 - \theta)^{N_I - x + \beta - 1}}{B(x + \alpha, N_I - x + \beta)} \Pr(N_D = x | N_I) \end{aligned} \tag{12}$$

Considering the N_I IDSes traversed by the intrusion flow, we can estimate the detection probability p as a prior in Eq. 5 taking the expected in Eq. 12.

$$
\begin{aligned}
E_{\Pr(\theta|N_I)}(\Theta) &= \int_0^1 \theta \Pr(\theta|N_I) d\theta \\
&= \int \sum_{x=0}^{N_I} \Pr(N_D = x|N_I) \frac{\theta^{x+\alpha}(1-\theta)^{N_I-x+\beta-1}}{B(x+\alpha, N_I-x+\beta)} \\
&= \sum_{x=0}^{N_I} \Pr(N_D = x|N_I) \frac{\int_0^1 \theta^{x+\alpha}(1-\theta)^{N_I-x+\beta-1} d\theta}{B(x+\alpha, N_I-x+\beta)} \\
&= \sum_{x=0}^{N_I} \Pr(N_D = x|N_I) \frac{B(x+\alpha+1, N_I-x+\beta)}{B(x+\alpha, N_I-x+\beta)} \\
&= \sum_{x=0}^{N_I} \Pr(N_D = x|N_I) \frac{x+\alpha}{N_I+\alpha+\beta} \\
&= \frac{1}{N_I+\alpha+\beta} \sum_{x=0}^{N_I} \Pr(N_D = x|N_I)(x+\alpha) \\
&= \frac{1}{N_I+\alpha+\beta} \left(\alpha + \sum_{x=0}^{N_I} \Pr(N_D = x|N_I) x \right) \\
&= \frac{\alpha + E(N_D)}{N_I+\alpha+\beta}
\end{aligned}
\tag{13}
$$

$E(N_D)$ can be evaluated as a fraction of IDSes traversed by a malicious flow (N_I) that detected an intrusion with probability PR_{av}.

$$
E(N_D) = N_I . PR_{av} \tag{14}
$$

In Equation 14, PR_{av} is just the mean result obtained from two dataset rounds of training over the same baseline IDS (Snort). What we need is to estimate the posterior probability $E_{\Pr(\theta|N_I)}(\Theta)$ that will replace the prior p in Eq. 5.

$$
p \approx E_{\Pr(\theta|N_I)}(\Theta) = \frac{\alpha + N_I . PR_{av}}{N_I+\alpha+\beta} \tag{15}
$$

Equation 15 updates the prior knowledge from a dataset training that measures the capacity of a supposed averaged IDS in the federation to detect either correctly or incorrectly an intrusion $p = \Pr(U_i = 1)$ with the current detection scenario, given by the number of detection messages received at the target (N_D) and the DIDS platform size (N_I).

4.1 Performance metrics

The true-positive rate of the DIDS platform (TPR_{DIDS}) measures the positive detection accuracy of the DIDS federation as a whole. In other words, the higher the TPR value, the bigger the number of detectable intrusions by the DIDS platform. As a posterior probability, the TPR_{DIDS} model proposed in [9] considers the mean positive-prediction value (PPV_{av}) of an IDS member as the belief mass of each arriving BGP message to be combined at the destination. In the same work, the prior probability, related to the previous knowledge about the DIDS capacity to detect (correct or incorrect) an intrusion, is obtained according to Eq. 1. That is, by averaging the confusion matrix from multiple dataset trainings in the same an standard IDS (Snort).

$$
\begin{aligned}
TPR_{DIDS} &= \frac{\Pr(U=1|I=1)}{\Pr(I=1)} \\
&\geq \left[1-(1-PPV_{av})^{N_D}\right] . \left[1-(1-PR_{av})^{N_I}\right]
\end{aligned}
\tag{16}
$$

Using the same probabilistic approach it is also possible to evaluate the false-negative rate (FNR) of the DIDS platform, which means the DIDS capacity to do not alarm normal traffic.

$$
\begin{aligned}
FNR_{DIDS} &= 1 - TPR_{DIDS} = \frac{\Pr(U=0|I=1)}{\Pr(I=1)} = \\
&\geq 1 - \left[1-(1-PPV_{av})^{N_D}\right] . \left[1-(1-PR_{av})^{N_I}\right]
\end{aligned}
\tag{17}
$$

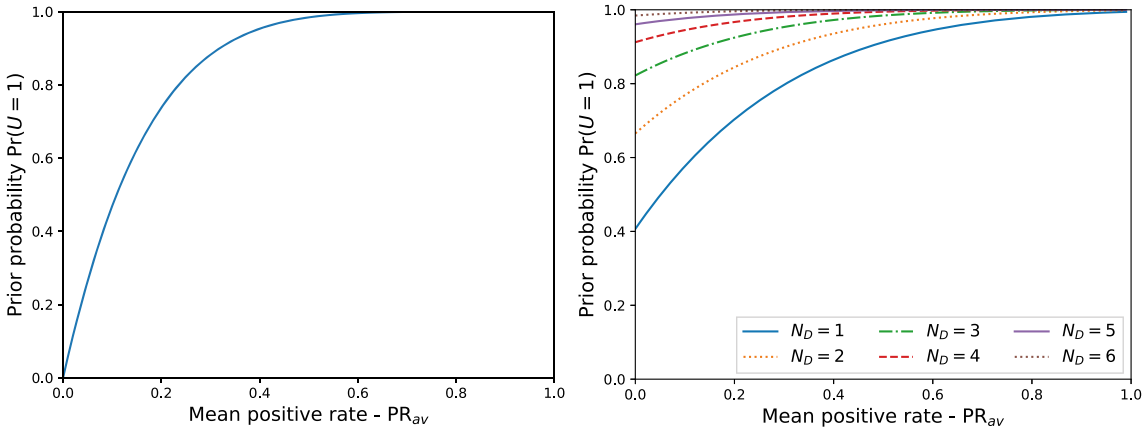
Equation 16 denotes TPR_{DIDS} as a posterior probability of the DIDS platform to detect an intrusion, given that there is a real intrusion occurring at some time[2]. The first right-side term is the likelihood, as the probability of having at least one correct detection among the N_D arriving messages from the IDS members. Likewise, the second right-side term denotes the flat empirical prior knowledge, as the probability of having at least one positive detection alarm (correct or incorrect) among the N_I IDS members traversed by the malicious flow. Using the balanced prior probability model proposed in Section 4, that is, updating PR_{av} with the intrusion evidence of the detection scenario, we can rewrite Eq. 16 using our prior estimation calculated in Eq. 15.

$$
\begin{aligned}
TPR_{DIDS} &= 1 - FNR_{DIDS} \\
&\geq \left[1-(1-PPV_{av})^{N_D}\right] . \left[1-(1-p)^{N_I}\right]
\end{aligned}
\tag{18}
$$

Likewise, we can rewrite the performance metric that measures the false-positive rate of the entire DIDS platform (FPR_{DIDS}) using the same prior probability proposed in Eq. 15.

$$
\begin{aligned}
FPR_{DIDS} &= 1 - TNR_{DIDS} \\
&\geq (1-PPV_{av})^{N_D} . \left[1-(1-p)^{N_I}\right]
\end{aligned}
\tag{19}
$$

[2]The denominator of the Bayes inference used to model the detection performance metric $\Pr(I=1)$ was assumed equals 1



(a) Flat prior probability model from [9] as a function of the mean positive rate PR_{av}.

(b) Balanced prior probability model from Section 4 as a function of the mean positive rate PR_{av}.

Fig. 3 Prior probability approaches to model the previous knowledge regarding the DIDS platform capacity to detect an intrusion $\Pr(U = 1)$. In this case, the value of PPV_{av} is fixed at 0.459. This number was obtained by averaging the confusion matrix presented in [28, 29]. **a** Flat prior probability model from [9] as a function of the mean positive rate PR_{av}. **b** Balanced prior probability model from Section 4 as a function of the mean positive rate PR_{av}

5 Comparative analysis

In this section, we compare the functional performance resulting from the models proposed in [9], which uses a flat empirical positive rate representing the previous knowledge about the detection probability, with the results obtained from the modeling approach presented in Section 4 that dynamically updates the prior probability with the numbers obtained in each detection scenario. Our goal is to analyze the main differences between the two modeling proposals and discuss their adherence with respect to the practical scenario described in [27], in which the number of intrusion messages from the sensing surface matters to evaluate the detection performance of the whole DIDS platform.

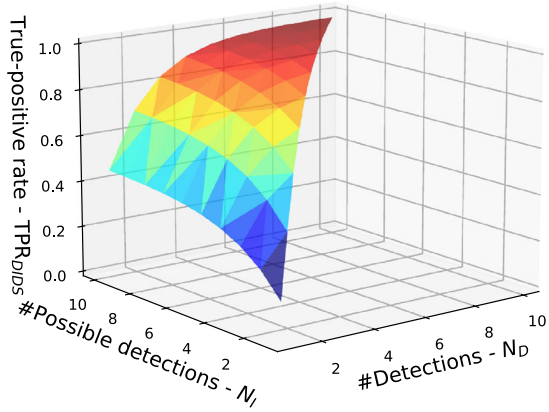
Figure 3 illustrates the behavior of the two prior probabilities models considered in this paper as a function of the mean positive detection rate of a standard IDS PR_{av}. As part of the Bayesian models for evaluating the functional performance of a DIDS, the prior probability denotes the previous knowledge regarding the DIDS capacity to detect an intrusion, which can be true or false. Figure 3a shows a single curve increasing exponentially with PR_{av}, as a result of the flat prior probability model proposed in [9] that does not consider the number of detection messages that arrive to be combined at the target AS N_D. On the other hand, Fig. 3b shows six different curves, according to the number of correlated BGP messages that can be combined at the destination AS N_D. In other words, the bigger the number of detection messages at the destination, the bigger the prior knowledge significance into the Bayesian metric model (posterior). The adaptable behavior for the prior probability proposed in this paper plays the role of balancing the detection metric, emphasizing the arriving data as intrusion evidence.

In order to ease the overall detection analysis, we focus our attention on the true-positive (TPR) and false-positive rates (FPR) which also grounds the evaluation of the remaining performance metrics, such as false-negative (FNR) and true-negative rates, respectively.

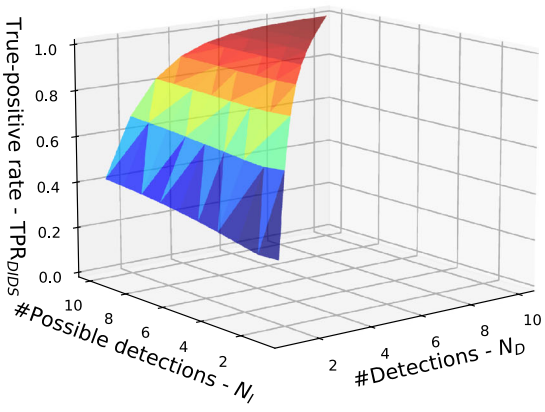
Figure 4 shows the TPR and FPR curves using the same input parameters. As one can see, regardless of the modeling approach, the detection metrics behave as expected, improving as both N_I and N_D increase. Figure 4a shows the TPR performance behavior using the flat model proposed in [9], where the prior knowledge regarding the positive accuracy of the DIDS platform is evaluated empirically, by averaging the positive rate (PR_{av}) from two dataset trainings [28, 29]. On the other hand, Fig. 4b shows the TPR curve using the balanced approach proposed in this paper, in which the prior knowledge is dynamically updated based on the detection numbers N_D and N_I.

As one can see, although the two TPR and FPR curves perform quite similarly among them, there are subtle differences regarding their behaviors, where the most remarkable one is when $N_D = 1$. In this case, both TPR (Fig. 4d) and FPR (Fig. 4c) are flatter than their respective curves shown in Fig. 4a and c. In other words, even being exponentially related to the federation scale (N_I), the detection performance does not increase too much because $N_D = 1$ means just a single piece of intrusion evidence at the destination AS.

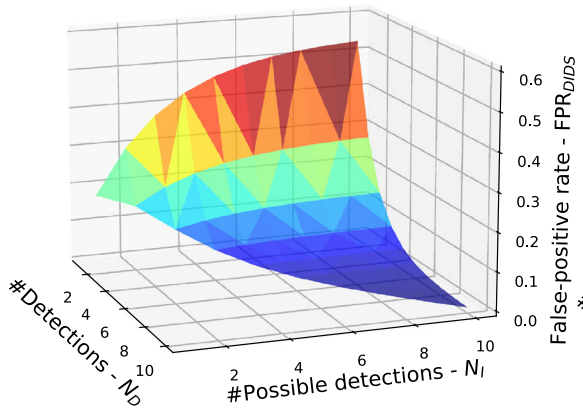
Figure 5 compares the TPR curves from the two models considered in this paper, as a function of PPV_{av}, for



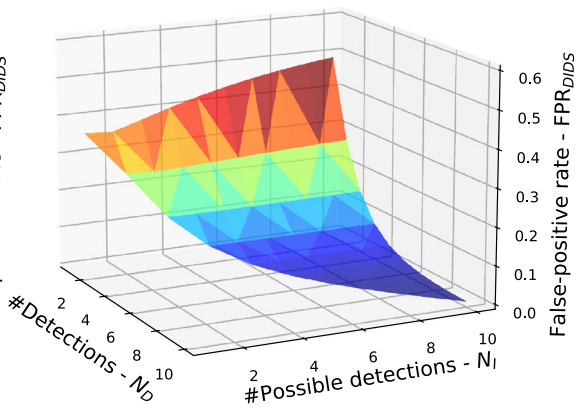
(a) TPR model using the flat prior proposed in [9].



(b) TPR model using the balanced prior proposed in Section 4.

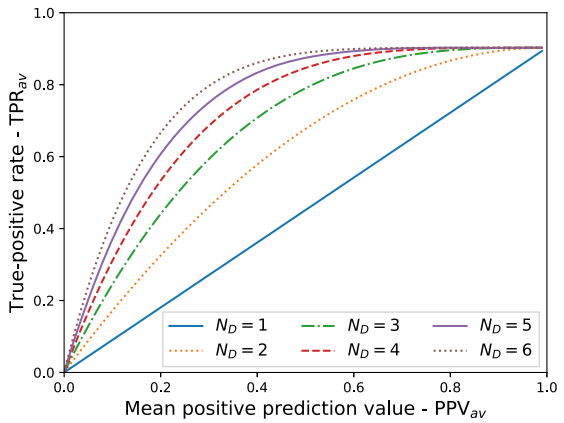


(c) FPR model using the flat prior proposed in [9].

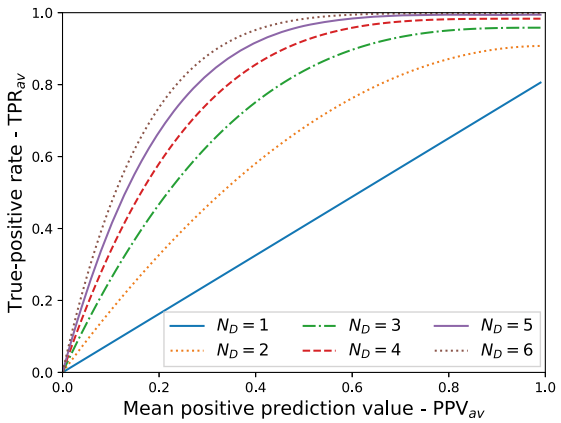


(d) FPR model using the balanced prior proposed in Section 4.

Fig. 4 True-positive and false-positive rates analysis of the overall DIDS platform (TPR$_{DIDS}$) as a function of the N_D and N_I. In this case, the values of PPV$_{av}$ and PR$_{av}$ are held at 0.459 and 0.322. These numbers were obtained averaging the confusion matrix presented in [28, 29]. **a** TPR model using the at prior proposed in [9]. **b** TPR model using the balanced prior proposed in Section 4. **c** FPR model using the at prior proposed in [9]. **d** FPR model using the balanced prior proposed in Section 4



(a) TPR metric, considering the flat prior model proposed in [9].



(b) TPR metric, considering the balanced model proposed in Section 4.

Fig. 5 True-positive rate analysis of the overall DIDS platform (TPR$_{DIDS}$) as a function of the PPV$_{av}$, for different values of and N_D. In this case, the PR$_{av}$ value is fixed at 0.322 [28, 29]. **a** TPR metric, considering the at prior model proposed in [9]. **b** TPR metric, considering the balanced model proposed in Section 4

Table 2 TPR and FPR results from the models proposed in [9] and Section 4 (holding $PR_a = 0.322$ and $PPV_{av} = 0.459$) for $N_D = Q = 6$ with the numerical results presented in IDN [16] and in [27] (practical testbed)

Metric	Model [9]	Model Section 4	IDN [16]	Testbed [27]
TPR	0.88	0.97	0.95	0.98
FPR	0.022	0.025	0.035	0.020

different values of N_D. As can be observed, the detection performance curves presented in Fig. 5a and b are quite similar and behave both like our expectations, that is, increasing fast as PPV_{av} and N_D also increase. However, while the TPR metric presented in Fig. 5a performs better for $N_D = [1, 3]$, Fig 5b outperforms the previous for $N_D = [4, 6]$.

The modeling framework proposed in this paper can also be used to assess the functional performance metrics of other collaborative DIDS approaches. For example, we can assume the number of acquaintance interactions among IDS agents Q from the intrusion detection network system (IDN) proposed in [16] as being the number of correlated messages to be combined N_D as in our analytic model. Table 2 puts together the numerical results obtained from the two models considered in this paper and the performance metrics from the IDN system proposed in [16] regarding the practical outcomes from the testbed presented in [27].

The numerical values arranged in Table 2, shows that the TPR and FPR outcomes from the model proposed in Section 4 of this paper are closer to the system proposed in [16] as well as to the practical numbers obtained from Testbed [27]. An additional assessment of the modeling performance of the proposed approach can be made by comparing the numerical results in Table 2 with Fig. 5a and b. While the balanced model tends to be one in the best detection situation, the previous flat model does not meet TPR values above 0.9 as the ones from IDN [16] and Testbed [27] depicted in Table 2. It can be explained by considering the dynamic behavior of the Beta function, which emphasizes the TPR improvement for positive detection situations $\alpha > \beta$.

6 Conclusion and future works

The main points that credence the DIDS approach as an evolution regarding the typical IDS architectures are (i) on improving the functional performance by extending the detection surface and exploiting detection consensus, and (ii) on eliminating a single point of failure related to the perimeter-based monolithic approach of typical IDSes.

Expanding the detection surface means increasing the number of federation members and encouraging their autonomy and heterogeneity. Likewise, exploiting detection consensus means using the multiple intrusion evidence from the federated agents to support a security decision, based on the belief degree of the combined information. In this regard, although the results presented in Section 5 show a lower performance for $N_D = [1, 3]$ concerning the previous modeling approach presented in [9], the results obtained from the proposed model, in which the data received from sensing agents balance the prior knowledge about the detection rate of the DIDS platform, are closer to the DIDS practical scenario shown in [27]. For further works, we plan to compare the performance evaluation results obtained from different analytic models with a practical DIDS platform, composed by the same number of IDS agents, preferably spread network-wide. It is also part of our plans to analyze the DIDS functional performance using federated learning approaches.

Acknowledgements The authors thank FAPERJ — the official funding agency for supporting science & technology research in the State of Rio de Janeiro (Brazil) and Rede-Rio (the state academic backbone network) — for the support given in the course of this work.

References

1. Möller DPF (2020) Intrusion detection and prevention. Springer International Publishing, Berlin, pp 47–75
2. Kumar Ahuja DG (2015) Evaluation metrics for intrusion detection systems-a study. Int J Comput Sci Mobile Appl 11:7–11
3. Ring M, Wunderlich S, Scheuring D, Landes D, Hotho A (2019) A survey of network-based intrusion detection data sets. Comput Secur 86:147–167
4. Gharib A, Sharafaldin I, Lashkari AH, Ghorbani AA (2016) An evaluation framework for intrusion detection dataset. In: 2016 International conference on information science and security (ICISS). 1–6
5. Sharafaldin I, Lashkari AH, Ghorbani AA (2018) Toward generating a new intrusion detection dataset and intrusion traffic characterization. In: Mori P, Furnell S, Camp O (eds) Proceedings of the 4th international conference on information systems security and privacy, ICISSP 2018, Funchal, Madeira - Portugal, January 22-24, 2018, SciTePress 108–116
6. Shah SAR, Issac B (2018) Performance comparison of intrusion detection systems and application of machine learning to snort system. Futur Gener Comput Syst 80:157–170
7. Robbins R (2003) Distributed intrusion detection systems: An introduction and review. Technical Report version 1.4b, ©SANS Institute Information Security Reading Room (February 2003) Option1
8. Jaynes ET (1968) Prior probabilities. IEEE Trans Syst Sci Cybern 4(3):227–241
9. Silva RS, de Moraes LF (2019) A cooperative approach with improved performance for a global intrusion detection systems for internet service providers. Ann Telecommun 74(3):167–173
10. Ghosh A, Sen S (2005) Agent-based distributed intrusion alert system. In: Sen A, Das N, Das SK, Sinha BP (eds) Distributed computing - IWDC 2004. Springer Berlin Heidelberg, Berlin, pp 240–251

11. Janakiraman R, Waldvogel M, Zhang Q (2003) Indra: a peer-to-peer approach to network intrusion detection and prevention. In: WET ICE 2003. Proceedings. Twelfth IEEE international workshops on enabling technologies: infrastructure for collaborative enterprises, 2003, pp 226–231
12. Cai M, Hwang K, Kwok Y-K, Song S, Chen Y (2005) Collaborative internet worm containment. IEEE Secur Privacy 3(3):25–33
13. Yegneswaran V, Barford P, Jha S (2004) Global intrusion detection in the domino overlay system. In: NDSS, The Internet Society
14. Zhou CV, Karunasekera S, Leckie C (2005) A peer-to-peer collaborative intrusion detection system. In: 2005 13th IEEE International conference on networks jointly held with the 2005 IEEE 7th malaysia international conf on communic, vol 1, p 6
15. Zaman S, Karray F (2009) Collaborative architecture for distributed intrusion detection system. In: 2009 IEEE symposium on computational intelligence for security and defense applications, pp 1–7
16. Fung CJ, Boutaba R (2013) Intrusion detection networks - a key to collaborative security. CRC Press
17. Zhao R, Yin Y, yu Shi Y, Xue Z (2020) Intelligent intrusion detection based on federated learning aided long short-term memory. Phys Commun 42:101157
18. Li K, Zhou H, Tu Z, Wang W, Zhang H (2020) Distributed network intrusion detection system in satellite-terrestrial integrated networks using federated learning. IEEE Access 8:214852–214865
19. Jamieson KG, Gupta MR, Krout DW (2009) Sequential bayesian estimation of the probability of detection for tracking
20. Shen Y, Cooper G (2010) A new prior for bayesian anomaly detection. Methods of Inf Med 49(1):44
21. Loibl C, Hares S, Raszuk R, McPherson B, Bacher M (2020) Dissemination of flow specification rules. RFC 8955 IETF©
22. Shenfield A, Day D, Ayesh A (2018) Intelligent intrusion detection systems using artificial neural networks. ICT Express 4(2):95–99
23. Vinayakumar R, Alazab M, Soman KP, Poornachandran P, Al-Nemrat A, Venkatraman S (2019) Deep learning approach for intelligent intrusion detection system. IEEE Access 7:41525–41550
24. Lynch SM (2007) Basics of Bayesian Statistics. Springer New York, New York, pp 47–75
25. Kak A (2017) Ml, map, and bayesian — the holy trinity of parameter estimation and data prediction. Tutorial, Purdue University©
26. Bolstad WM, Curran JM (2007) 8 Bayesian Inference for Binomial Proportion. Wiley, New York, pp 141-159
27. D'Antonio S, Formicola V, Mazzariello C, Oliviero F, Romano SP (2010) Performance assessment of a distributed intrusion detection system in a real network scenario. In: 2010 Fifth international conference on risks and security of internet and systems (CRiSIS), pp 1–8
28. Najafian Z, Aghazarian V, Hedayati A (2015) Signature-based method and stream data mining technique performance evaluation for security and intrusion detection in advanced metering infrastructures (ami). Int J Comput Electr Eng 7:128–139
29. Valero León A (2017) INsIDES: A new machine learning-based intrusion detection system. Ph.D. dissertation, Universitat Pompeo Fabra, Barcelona

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.