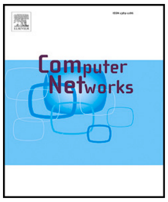
Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet



Online Distributed Denial of Service (DDoS) intrusion detection based on adaptive sliding window and morphological fractal dimension

Gianmarco Baldini [a,*], Irene Amerini [b]

[a] European Commission, Joint Research Centre, Ispra, Italy
[b] Sapienza University of Rome, Department of Computer, Control and Management Engineering Rome A.Ruberti, Italy

ARTICLE INFO

Keywords:
Intrusion detection systems
Security
Fractals
Machine learning
Communication

ABSTRACT

Distributed Denial of Service (DDOS) attacks are important threats to network services and applications. Studies in literature have proposed various approaches including Intrusion Detection Systems (IDS) based on the application of machine learning and deep learning, but their computational cost can be significant. For this reason, other studies have proposed efficient IDS algorithms based on the online real-time analysis of the network traffic with a sliding window and entropy or other statistical measures. This paper proposes an online algorithm based on a sliding window with the novel application of the Morphological Fractal Dimension (MFD) to this problem. The results presented in this study show that the application of MFD to the recent CICIDS2017 public data set can obtain a significant improvement in the detection of the DDoS attack in comparison to entropy based approaches. In addition, this paper proposes a novel algorithm for the automatic definition of the sliding window size. This paper reports the impact of the different hyper-parameters, including the parameters present in the definition of MFD and the evaluation of the distance measures, where the Chebyschev distance provides the optimal detection accuracy. The results show a detection accuracy of 99.30%, which performs better than similar approaches on the same data set.

1. Introduction

Society and the citizens are becoming increasingly dependent on the internet and communication services but the risk of cybersecurity threats has also increased. One of the most significant cybersecurity attacks is the Denial-of-service (DoS) attack, where legitimate users are prevented from accessing a network service because the attacks overwhelm the network resources. Distributed DoS (DDoS) is a sophisticated and distributed many-to-one version of the DoS attack, which can be even more complex to detect and mitigate [1]. Cybersecurity attacks and more specifically DDoS can be mitigated by Intrusion Detection Systems (IDS), designed to detect traffic anomalies associated with the implementation of an attack. Requirements or preferred features of intrusion detection systems have been already defined in the literature [2,3] and they can be summarized as following: (a) fast detection of the attack, (b) high detection accuracy, (c) low computing complexity of the detection algorithm to support the capability to analyze a large amount of traffic due to the high throughput of the current networks. The successful fulfillment of these three main requirements (and other requirements defined in the literature) can be challenging because there are trade-offs among the requirements. For example, algorithms that can obtain high detection accuracy may require considerable computing resources or they may not be able to achieve a fast detection. This paper focuses on the optimization of the requirements for high detection accuracy and fast detection using a sliding window approach where a window of traffic data is analyzed and compared against a predefined model. Then, the data dimensionality reduction is implemented by feature extraction on a non-overlapping sliding window, where each window collects a set traffic flows statistics on which high level features are calculated. In general, feature extraction is the process of accurately simplifying the representation of data (e.g., the network traffic statistics in this context) by reducing its dimensionality while extracting its relevant characteristics for the desired task of attack detection. Feature extraction has a substantial effect on the attack detection accuracy and speed since classification carried out without a successful feature extraction process on a high dimensional and redundant data would be computationally complex and it may over-fit the training data.

Many algorithms proposed in literature (see Literature Review Section 2) adopt a dimensionality reduction of the data traffic where traffic features like the number of packets in the flow are extracted from the network flows. An additional step is to apply specific transforms on the traffic features of a specific segment of the traffic data (i.e., the

[*] Corresponding author.
E-mail address: gianmarco.baldini@ec.europa.eu (G. Baldini).

https://doi.org/10.1016/j.comnet.2022.108923
Received 16 February 2021; Received in revised form 3 January 2022; Accepted 15 March 2022
Available online 5 April 2022

sliding window). Entropy-based detection algorithms (i.e., application of entropy measures) have been used in literature in this approach because it has been proven that network attacks change the entropy of the network traffic statistics, in particular in the case of the DDoS attack [1,4]. Various entropy measures have been adopted in literature including Shannon, Renyi, and Tsallis entropy as described in Section 2. This paper proposes a novel method to intrusion detection with a sliding window approach based on the Morphological Fractal Dimension (MFD), which has been introduced in [5] and which has not been applied to this specific context to the best knowledge of the authors.

In particular, this paper provides the following main contributions:

1. While in literature, statistical features (e.g., variance) and entropy measures (e.g., Shannon entropy) are calculated on the sliding window traffic flow statistics, this paper proposes the application of the MFD in one dimension, which has been introduced in [5]. The concept of morphological transformations (e.g., dilation and erosions) are well known in image processing together with the estimate of the corresponding fractal dimensions, but the application of these concepts to cybersecurity and IDS is novel to the best knowledge of the authors. The underlying concept for the application of MFD to the IDS context is that attacks can generate anomalies in the fractal complexity of the traffic statistics, which can be exploited to detect the attacks. In particular, the appropriate (in terms of hyper-parameters selection) application of morphological dilation and erasures is similar to the application of entropy measures to the traffic statistics which has similar discriminating power (i.e., because the entropy of the traffic statistics changes when the attack is implemented). The results of this study show that the application of MFD has more discriminating power (which translates to a higher attack detection accuracy) than the entropy measures applied in the literature. The trade-off is the need to identify the optimal values of the hyper-parameters defined in the design of MFD.
2. This paper proposes a novel adaptive sliding window method for IDS where the optimal (i.e., optimal for the accuracy of the attack detection) size of the sliding window is calculated adaptively. In particular, the goal is to define a time efficient adaptive method which does not require the execution of the machine learning classification algorithm but which only relies on the statistical properties of the network statistics to define the optimal value of the sliding window. The results from the application of the adaptive approach are compared with the ground truth where the detection accuracy is evaluated for different sizes of the sliding window.

The authors would also like to highlight that the approach is evaluated using the recent public data set CICIDS2017 rather than the outdated DARPA KDD-99 data set, which is still widely used in academic research [6].

The structure of this paper is the following: Section 2 provides the literature review. Section 3 describes the online sliding window approach adopted in this paper with its main workflow of operations, the definition of MFD, the evaluation metrics and the materials (i.e., CICIDS2017 data set) used to evaluate the approach. Section 4 presents the results, including the findings from the hyper-parameters optimization phase and the comparison to the other approaches used in the literature. Finally Section 5 concludes this paper.

2. Literature review

The concept of Intrusion Detection System (IDS) has a long history in literature as described in one of the initial surveys on this topic in 1993 [3], where it is mentioned that IDS performs the essential function to detect unauthorized intruders and attacks to the network infrastructure. These attacks are possible because computer and network infrastructures are often deployed in the market with security vulnerabilities, which are still present since the goal of building systems with no vulnerabilities is extremely difficult if not possible [3]. Then, monitoring systems like IDS, which analyzes traffic logs and audit trails are useful to detect attacks, which exploit vulnerabilities after the market deployment. The survey [3] mentions two main IDS categories: offline IDS where the analysis of logs and audit records is performed some time after the traffic network operation (e.g., the analysis is executed the day after the network or computer system activity) and the online (or real-time) IDS where the analysis is performed directly on the traffic or immediately after the traffic features are calculated (e.g., average duration of the packets or average time of the connection). For example, the online IDS performs the analysis on a single or a set of observations (e.g., network flows) at the time after an initial training phase, while the offline IDS analyzes all the observations of the day before. The IDS belonging to the second category (i.e., online) would be preferable because the attack can be quickly detected and acted upon by a network manager or system administrator before the attack can impact negatively the network infrastructure and cause additional damage. On the other side, online IDSs are more difficult to implement especially in current networks infrastructures because they must be very fast but still reasonably accurate. Regarding the terminology, we adopt in this paper the same understanding of online and real-time proposed in the 1993 survey [1] where it is recognized that the terms "online" and "real-time" tend to be used interchangeably in the literature, but we highlight the difference that an ideal online machine learning algorithm should be able to update itself and adapt to a frequently changing environment.

From the initial survey of 1993 [1], the literature presents hundreds of studies implementing both offline and online IDS. Since this paper proposes an online IDS, we focus the literature on this specific category, which can be further divided into sub-categories. More recent surveys like [7,8] provide different taxonomies for online IDS and the approach proposed in this paper can be classified in the category of anomaly detection where the activities of a system at an instant (e.g., an observation or set of observations) are compared against the normal behavior profile calculated in a training phase against legitimate traffic. The IDS generates the alert whenever a deviation from normal behavior exceeds a threshold. This approach is quite efficient to detect new attacks, in particular, those attacks related to the abuse of computing resources, but any observation, which does not match to a normal behavior is considered an intrusion and this may lead to poor accuracy [8]. Researchers have used machine learning algorithms approaches, statistical based approaches or information theory based approaches. The latter category (to which the approach proposed in this paper belongs) includes a number of recent studies where entropy measures are used to model the normal behavior and the attacks on the network on the basis of the hypothesis (proven in the studies) that the entropy of the traffic changes when an attack is implemented. In particular, the authors in [9] propose a detection method called D-FACE to differentiate legitimate and DDoS attacks. The method compares the source IP Shannon entropy of the normal traffic flows with the traffic in a specific time window (e.g., the observation). This entropy difference is called Information Distance (ID) and is used as the detection metric when the calculated entropy goes beyond the thresholds associated with legitimate traffic. Our approach adopts a similar approach based on a sliding window and thresholds with the difference that the sliding window is adaptive and the MFD is used instead of the Shannon entropy. In addition, our approach adopts other features in addition to the source IP. A paper, which also uses network flow features (e.g., Shannon entropy) to detect DDoS as in our approach is [10]. In similar way, the authors in [11] use fast entropy to detect DDoS attacks as they demonstrate that the flow count entropy severely decreases in the case of attack flows, and it is stable otherwise. A pre-defined threshold is also used in this case and the difference between the flow count entropy and the mean entropy of the legitimate traffic is used to detect the DDoS attack. A sophisticated approach to evaluate the difference between legitimate traffic and anomalous traffic potentially linked to DDoS attack is presented in [1] where Shannon entropy (together with statistical features) are extracted from network flows per minute as detection metrics. Then, the authors employ a kernel-based learning algorithm (i.e., KOAD) using the entropy features to detect input vectors that were suspected to be DDoS. The study presented in this paper has various similarities to [1] since it uses the same data set: the CICIDS2017 dataset, the same concept of aggregating a set of network flows for efficiency, the selection of specific network flows features and the use of distance metric for detection. The authors in [1] use Shannon entropy. On one side, this paper proposes a simpler detection approach more aligned to the other studies [10,11]. On the other side, this paper proposes the use of MFD which is demonstrated to be more accurate in this paper than the Shannon entropy. Still, the sophisticated approach proposed in [1] based on KOAD demonstrates superior performance to the basic detection approach and future developments of this paper could integrate the MFD with KOAD.

All the papers cited above have used Shannon entropy as high-level feature based on a fixed sliding window. Other studies have used different entropy measures. The authors in [4] have investigated and compared Shannon entropy, Renyi entropy, and Tsallis entropy for various attacks including DDoS. The results show that for the specific data set used in [4], Renyi entropy and Tsallis entropy provide a higher detection accuracy than Shannon entropy. The authors in [12] have also used Renyi entropy and Tsallis entropy in addition to Shannon entropy to detect attacks, but no fractal dimensions were used in the analysis. The authors of [13] have investigated the impact of the α parameter in the application of Renyi entropy for the detection of low-rate DDoS attacks. The authors of [14] have proposed an anomaly detection method to detect DDoS attacks based on the combination of Tsallis Entropy and the Lyapunov exponent. The approach is based on the comparison of the entropy between Source IPs and Destination IPs, by analyzing the rate of exponent separation. Also, in this case, no fractal dimension was used.

Very few papers have used fractal dimensions to implement IDS. The authors in [15] have used fractal dimension D and Hurst parameter H to detect the DDoS attack on the DARPA intrusion detection evaluation data sets but they have not used the MFD. In addition, the DARPA intrusion detection evaluation data sets are now considered outdated [7]. The author of [16] has instead used the Higuchi fractal dimension for intrusion detection systems, where it has demonstrated its capability to detect an attack, but no comparison with other features was provided. In [17] the authors have used Hausdorff fractal dimension to detect Denial Of Service (DOS) attacks in combination with Shannon entropy measure on the probabilities of time series of measured ICMP request/echo delays extracted from the network traffic.

In comparison to the state of art described in the previous paragraphs, this paper describes a novel approach to IDS based on the MFD, which has not been attempted in literature yet. The MFD combines the aspects of morphological operations and fractal theory to detect changes in the structure of the traffic statistics. In addition, most of the identified papers used a fixed sliding window approach. This paper proposes an adaptive sliding window approach where the optimal size of the window is identified in a pre-processing step before the machine learning step, thus it is more computationally efficient than adaptive approaches which require the computation of the machine learning step. Finally, we would like to highlight that this approach is applied to a recent IDS data set while many papers cited above use the old and relatively obsolete DARPA KDD-99 dataset [6].

3. Methodology and materials

3.1. Workflow

The methodology is consistent with similar approaches adopted in other studies [1,4,18] and it is described in Fig. 1 where two phases are identified. In the first phase, a model of the legitimate traffic is created for each feature used in the analysis. The model is created by calculating the mean and standard deviation for each feature on the portion of the data set labeled as legitimate traffic. In the second phase, the traffic is analyzed in comparison to the legitimate traffic model. If there is a significant deviation, the traffic can be considered as an anomaly and it can be associated with an attack.

More in detail, the following steps are executed:

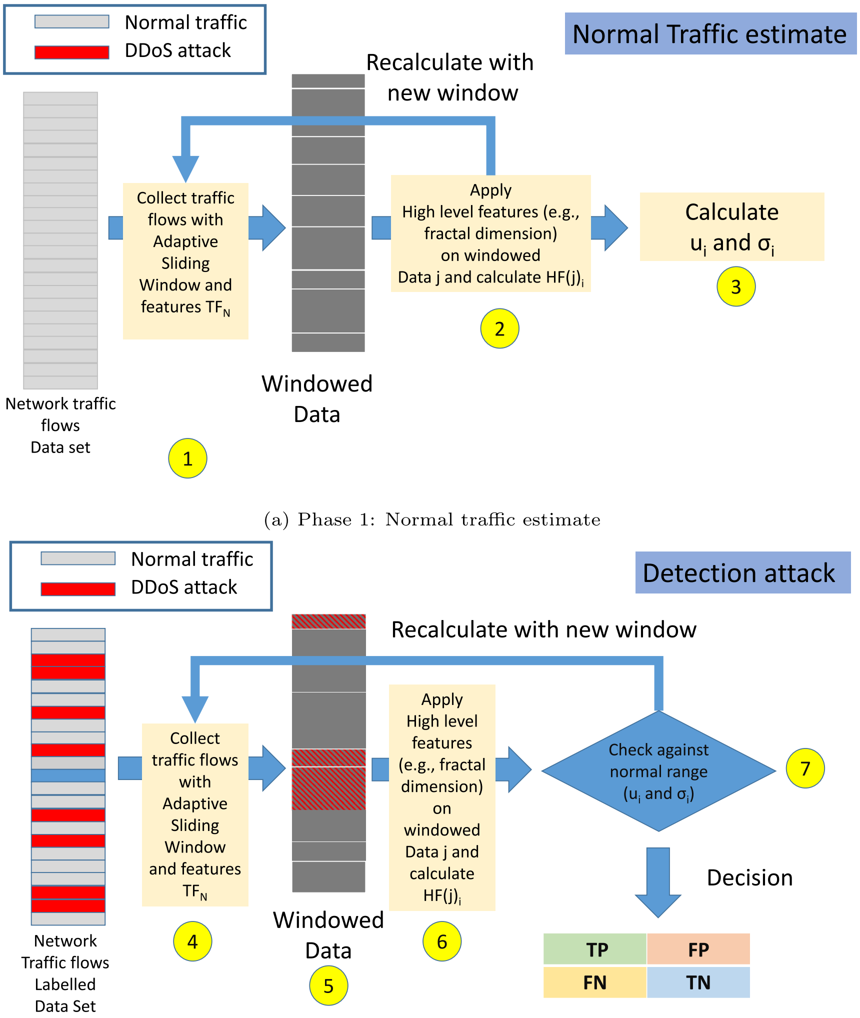
1. In the *Normal Traffic Estimate* phase, the network flows for the labeled legitimate traffic are collected in a sliding window (the windows are not overlapping) of size W_S for each Traffic Feature (TF) i with $i = 1..N$ (i.e., $N = 9$ in this study) (note that in the CICIDS2017 data set 78 network features are present as described in Section 3.5). The approach proposed in this paper is based on an adaptive determination of the value W_S using a specific function. This is described in Section 3.4.
2. High level Features (HF) (i.e., entropy or fractal dimension) are applied to the sliding window of size W_s and TF_i. Each HF is applied to each TF: $HF(j)_i$ with $j = 1..M$ (i.e., $M = 11$ in this study) and $i = 1..N$ (i.e., $N = 9$ in this study). The HFs used in this study are described in detail in Section 3.2.
3. The mean μ and standard deviation σ are calculated for each HF and TF (i.e., i, j) to the traffic network flows of the labeled legitimate traffic.
4. In the *Detection* phase, the TFs are extracted from the network traffic data as in the previous phase.
5. For each TF, $i = 1..N$, the sliding window of size W_S is calculated using the adaptive approach as in the previous phase.
6. $HF(j)_i$ (with $j = 1..M$) are calculated for each new sliding window of size W_s for each TF.
7. The new values of HF are compared against $\mu_{i,j}$, $\sigma_{i,j}$ of the legitimate traffic using different distance metrics. In this study, the Chebyschev distance provided optimal detection results (as shown in Section 4). Then, a decision is made against the labeled traffic with a threshold based on $\sigma_{i,j}$. A multiplier factor M_F for the threshold is used with threshold $= M_F * \sigma_{i,j}$. The range of the threshold is from 0.1 to 4 (1 means that the threshold is equal to the standard deviation). Then, the maximum value of the accuracy is selected in the threshold range.

We would like to highlight that the proposed approach does not negatively impact the flow of the benign traffic. This is a risk that other IDS approaches (e.g., based on traffic filtering) may incur as described in the survey papers [7,8], which describe the various IDS approaches. From this point of view, the approach proposed in this paper is passive because it is based on operations of feature extraction (e.g., MFD) on the network traffic statistics and machine learning classification, which can be implemented in computing platforms separated from the traffic layer. For example, this approach can be implemented in the network management layer, which is often responsible for the calculation of traffic statistics (e.g., for performance or billing purposes). Even if the proposed approach can exploit the existing network management functions for the analysis of traffic statistics, additional computing resources would be needed for the implementation of this approach.

3.2. High level features

In this paper, the following high level features HF are used: the Shannon entropy, the Renyi entropy of order 2, 3 and 4, the Tsallis entropy of order 2, 3 and 4, the MFD and the Katz, Higuchi and Hausdorff fractal dimensions. These high level features are used because: (a) they have already demonstrated their detection capabilities in IDSs proposed in literature and (b) to make a comparison among different fractal dimensions measures.

Regarding the application of entropy measures for IDS, the Shannon, Renyi and Tsallis entropy have been adopted for a comparison to


Normal traffic
DDoS attack
Normal Traffic estimate
Recalculate with new window
Network traffic flows Data set
Collect traffic flows with Adaptive Sliding Window and features TF_N
Windowed Data
Apply High level features (e.g., fractal dimension) on windowed Data j and calculate $HF(j)_i$
Calculate u_i and σ_i
1
2
3
Normal traffic
DDoS attack
Detection attack
Recalculate with new window
Network Traffic flows Labelled Data Set
Collect traffic flows with Adaptive Sliding Window and features TF_N
Windowed Data
Apply High level features (e.g., fractal dimension) on windowed Data j and calculate $HF(j)_i$
Check against normal range (u_i and σ_i)
Decision
TP
FP
FN
TN
4
5
6
7

(a) Phase 1: Normal traffic estimate

(b) Phase 2: Detection phase

Fig. 1. Methodology workflow with the two phases.

studies in literature using the same approach of the sliding window in [1,4] and [9,14] and [10]. The concept is that with cyber-security attacks and more specifically with DDoS attacks, the distributions of the network statistics will become more dispersed [14] and entropy measures could capture this dispersal phenomenon. Shannon entropy is the most common entropy measure used for this purpose and it was chosen for these reasons. Renyi entropy, and Tsallis entropy are generalizations of Shannon entropy and they could also indicate such dispersal of the distributions of the network statistics.

Regarding the fractal based features (i.e., Katz, Higuchi and Hausdorff fractal dimensions), they have been adopted as a comparison with the MFD and also because they have been recently proposed in IDS or to detect anomalies. The underlying concept of the application of fractal based features to the analysis of the network traffic is based on the consideration that the complexity of the network traffic statistics collected in a sliding window can change in time when an attack is being implemented. Because a fractal dimension is a ratio providing a statistical index of the complexity of the time series and/or it is a quantity that is self-similar over some region of space or time interval, it can be used to detect attacks. Even if this study is mainly based on MFD, other fractal dimensions (e.g., Higuchi) have been applied to make a comparison of the DDoS detection performance of MFD with other fractal dimensions.

For example, a recent paper [16] has used the Higuchi fractal dimension for IDS in [16], where it has demonstrated its capability to detect an attack, but no comparison with other features was provided. The authors have instead used the Hausdorff dimension in [17]. Note that the information provided above is only a summary of what is already reported in the Literature Review Section 2.

Table 1 summarizes the high level features used in this study for the DDoS attack detection together with the references of the used coding implementation and the list of studies using the specific features in a similar context (e.g., IDS).

Because MFD is the novel aspect of the approach proposed in this paper, it is described in detail in the next Section 3.3.

3.3. Morphological fractal dimension

Fractal are dimensional sets with a high degree of geometrical complexity, which can model many phenomena both natural (e.g., mountains or clouds as demonstrated in the seminal study by Mandelbrot in [24]) and/or related to human activities (e.g., urban agglomerations in [25]). While most of the studies have focused on the application of fractals for imaging, there are also many one-dimensional examples where a time series can be modeled as a fractal with the meaning

Table 1
List of high level features used for DDoS detection with related hyper-parameters and reference implementation.

High level feature	Hyper-parameter	Reference implementation	Reference studies
Shannon Entropy	None	MATLAB implementation (*entropy* function) from Mathworks.	[1], [4], [9], [10]
Renyi Entropy	Order p (p=2,3,4)	MATLAB implementation by [19]	[4] [12], [13],[20]
Tsallis Entropy	Order p (p=2,3,4)	MATLAB implementation by [19]	[4], [12], [14]
Morphological Fractal dimension (MFD)	g,c, k_{max}	MATLAB implementation provided in [21] based on an algorithm from [5].	None
Higuchi fractal dimension (HiFD)	Maximum number of sub-series	MATLAB implementation by [22]	[15],[16]
Katz fractal dimension (KFD)	Maximum number of sub series	MATLAB implementation by [22]	None
Hausdorff fractal dimension (HaFD)	None	MATLAB implementation by [23]	[17]

that their graph is a fractal set [5]. There are different definitions of fractals, but one of the most commonly adopted is the one by Mandelbrot in [24], where fractal is defined "a rough or fragmented geometric shape that can be split into parts, each of which is (at least approximately) a reduced-size copy of the whole". An important characteristic of fractals, which is used to classify and describe them, is the fractal dimension, which measures the degree of the boundary fragmentation or irregularity of time series (if applied to one dimension time series) over multiple scales. It makes meaningful the measurement of metric aspects of fractal curves, such as their length. As described in [5], given a measure unit of length E (i.e., the reference unit or 'yardstick'), the length L(E) of a curve at scale E is equal to the number of yardsticks that can fit sequentially along the curve times E.

Fractal dimension is able to describe the complexity degree of fractal set quantitatively, but it can also be applied to the evaluation of the complexity of a time series and signals. Among various methods to calculate fractal dimensions, fractal box counting dimension has a wide range of applications. The box counting method analyzes complex patterns by breaking a graph of the time series into smaller and smaller pieces, typically box-shaped (i.e., through a partition grid), and analyzing the pieces at each smaller scale. Some research [26] has shown that the calculated value is not precise due to the approximations of the partition grid in the estimate process. In addition, covering the graph of one-dimensional continuous-time signals by such planar sets (e.g., the boxes) involves two-dimensional processing of the signal at multiple scale with complexity $O(N^2)$. As demonstrated in [5], the MFD approach has instead $O(N)$ complexity, by using one-dimensional morphological filtering. In addition, it can yield results that are invariant with respect to shifting the signal's domain and/or affine scaling of its dynamic range. For these reasons, in this paper, we use the formulation of MFD as proposed in [5]. The main concept of MFD is to combine morphological covering method with estimates of the fractal dimensions. The process starts from an initial morphological estimate $D*$ of the time series f under study and the corresponding fractal function $f*$ is synthesized. Then, by searching in the parameter space D, a nonlinear optimization problem is performed, where a distance is iteratively minimized between the original f and each new iteratively synthesized $f*$. The process terminates when a minimum is reached. More in detail, the calculation method of MFD is described as follows: If we suppose that $f(i), i = 1, 2, \dots, N$ is a discrete signal in the time domain, g is the unit structuring element, k is the analytical scale, the structuring element with k scale is defined as follow (the k fold dilation of g with itself):

$$g_k \overset{\Delta}{=} \underbrace{g \oplus g \oplus g \oplus g \oplus g}_{(k-1)times} \tag{1}$$

where the function $\overset{\Delta}{=}$ is defined in [5] as $g(n) \overset{\Delta}{=} sup\{y : (n, y)\} \in B$ and $B \subseteq R^2$ is a compact set that is also single connected (i.e., connected with no holes) and symmetric with respect to both axes of the plane. The support-limited dilation and erosion of f by g with respect to a support set $S \subseteq R$ are defined for discrete time signals as (equations (33) and (34) of [5]):

$$f \oplus_S g_k = ((f \oplus_S g)\underbrace{\oplus_S g \cdots)\oplus_S g}_{(k)times} \tag{2}$$

$$f \ominus_S g_k = ((f \ominus_S g)\underbrace{\ominus_S g \cdots)\ominus_S g}_{(k)times} \tag{3}$$

One-dimensional morphological cover area with k scale is defined in the following equation:

$$A_g(k) = \sum_{(n=0)}^{N} ((f \oplus_S g_k)(n) - (f \ominus_S g_k)(n)) \tag{4}$$

with

$$k = 1, 2, 3, \dots, k_{max} < \frac{N}{2} \tag{5}$$

The fractal dimension D_M used in this study is the Minkowski–Bouligand dimension, which is conceptually based on Minkowski's idea of finding the length of irregular curves: dilate them with disks by forming the union of these disks centered at all points of n and thus create a Minkowski cover as described in [5]. The cover area $A_g(k)$ is able to meet the condition in which D_M is the MFD of the signal $f(n)$, c is a constant, k_{max} is the maximum value of k as defined in the following equation (which can be used to calculate D_M):

$$ln\left(\frac{A_g(k)}{k^2}\right) = D_M ln\left(\frac{1}{k}\right) + c, (k = 1, 2, \dots, k_{max}) \tag{6}$$

On the basis of the previous equations, there are three hyper-parameters to tune in the application of MFD to this particular problem: (1) the g unit, (2) the constant c (usually in a range of 0.001 and 0.0001 according to [5]) and (3) k_{max}. The g unit has a value of 1 for the segment, value of 2 for the square and value of 3 for the rhombus. While the parameter g and c are independent from the size of the sliding window W_s, the k_{max} parameter is related to W_S. Then, the definition of k_{max} must be related to W_S and a new parameter size factor S_F is introduced and the relation to k_{max} and W_S is defined by the following simple equation:

$$k_{max} = S_F \times W_S \tag{7}$$

A fourth hyper-parameter D_{dec} is the distance metric evaluated to compare the mean of the legitimate traffic model with the analysis

Table 2
List of measures used in the adaptive window size calculation.

Measure	Hyper-parameter	Reference implementation
Variance	None	MATLAB implementation (*var* function) from Mathworks.
Burstiness Statistic	None	MATLAB implementation by [19] of the algorithm described in [28].
Shannon Entropy	None	MATLAB implementation *entropy* function) from Mathworks.
Quantile	Cumulative probability threshold p	MATLAB implementation *quantile* function from Mathworks.

of the traffic data. Various distance metrics are evaluated: Manhattan distance, Euclidean distance, Minkowski distance of order $o = 3, 4, 5$ and Chebyschev distance. Note that Manhattan distance and Euclidean distance are respectively Minkowski distances of order $o = 1, 2$ respectively.

The impact of these hyper-parameters on the DDoS detection performance is evaluated and presented in Section 4.

3.4. Adaptive sliding window

This paper proposes a sliding window method where the segment size is obtained in a dynamic way to optimize the detection accuracy. This approach is not novel in time series analysis and classification where various adaptive sliding window approaches are proposed, but it is not proposed in IDS systems so far to the best knowledge of the authors. The challenge is to identify a statistical measure, which is both time efficient and which provides a high accuracy at the same time. In literature, the standard deviation is used in combination with sliding windows while other methods use metaheuristic algorithms for search optimization (e.g., genetic algorithm) in combination of machine learning algorithm which can be quite time consuming. For example, the authors in [27] propose an approach where the size of the window is calculated using an optimization method based on deep learning, which can be demanding from the computing point of view. In this paper, we use an approach based on a very simple statistical measure. This study has applied different measures described in Table 2, where the link to the MATLAB implementation is also provided for reproducibility of the results. The performance of each algorithm in terms of accuracy and processing time is shown in Section 4. The algorithm for the adaptive selection of the sliding window is implemented as follows. In the first step, the feature described in the first column of Table 2 is applied to a range of window sizes W_S which is bounded from the lower side (e.g., smallest segment size) by the minimum time series length required by the application of MFD and from the high side (e.g., largest segment size) to provide a number of samples large enough to be meaningful from the statistical point of view for the training and detection phase. On each of the window sizes, the measures described in Table 2 are calculated. Then, the minimum value of the calculated measures across all the segment sizes is used to indicate the proper segment size. The reason why the minimum value is chosen is related to the following considerations. Firstly, the results of the empirical evaluations show that the minimum value provides the highest overall detection accuracy. Secondly, a smoother segment (i.e., less bursty) is more aligned to the specific type of traffic in the observation (legitimate or attack) and therefore the segment is more representative. Then, the segment size associated with the calculated minimum value is used to extract the High level Features (HF). Various measures identified in 2 are evaluated for their performance in terms of accuracy and time with the Quartile measure with cumulative probability threshold p achieving the optimal performance as demonstrated in Section 4 (i.e., Table 4).

3.5. Materials

To evaluate the proposed approach, the publicly available CICIDS2017 data set provided in [29] is used. The dataset contains normal traffic (i.e., legitimate traffic with no attacks) and traffic with the most up-to-date common attacks for five days. The data of July 03, 2017 includes normal traffic, and the data of July 07, 2017 includes both normal and DDoS attack traffic. The DDoS traffic in this dataset was generated with a tool to flood UDP and TCP requests to simulate network layer DDoS attacks, and HTTP requests to simulate application-layer DDoS attacks. The dataset is completely labeled and includes 78 network traffic features, which were extracted using the CICFlowMeter software package described in [29]. The TF features of this vector were determined based on the list of the best 9 selected features for DDoS attack detection and the legitimate traffic identified in [29]: Port Id, Minimal length of the backwarding packets, Subflow Forwarding bytes, Total length of Forwarding packets, Mean length of Forwarding packets, Standard deviation of the Backward packet length, Average Packet size, Duration of the Flow and Standard deviation of the Flow Inter arrival time. The identification of these optimal features was also assessed by the authors of this study with an independent evaluation which confirmed the results of [29]. A similar result was also achieved by other authors using the same data set [1]. This assessment is not shown in this paper for reasons of space.

The portion of the CICIDS2017 data set of July 03, 2017, which includes normal traffic was used for the phase 1 of the methodology described in Fig. 1, while the rest of the data set was used for the phase 2 of the methodology. This is a similar approach adopted by researchers using the CICIDS2017 data set [1,29].

3.6. Evaluation metrics

This subsection describes the metrics used to evaluate the performance of the approach proposed in this paper and the alternative approaches used in the literature.

The main metric is the detection accuracy, which is defined as:

$$Accuracy = \frac{TP + TN}{(TP + FP + FN + TN)} \tag{8}$$

Where TP is the number of True Positives, TN is the number of True Negatives, FP is the number of False Positives and FN is the number of False Negatives.

To complement the accuracy metric, the True Positive Rate (TPR) (also called sensitivity), the precision and the False Positive Rate (FPR) (fall-out) are used, which are defined in the following equations:

$$TPR = \frac{TP}{(TP + FN)} \tag{9}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{10}$$

$$FPR = \frac{FP}{(FP + TN)} \tag{11}$$

where the TP is the number of True Positives (legitimate traffic predicted as such), TN is the number of True Negatives (attack-related traffic predicted as such), FP is the number of False Positives (attack-related traffic predicted as legitimate traffic) and FN is the number of False Negatives (legitimate traffic predicted as attack-related traffic).

An additional metric used in this study is the time needed to detect the attack (detection time or T_D), which is related to the computational complexity of the approach. Because the proposed approach is based on a number of different steps (calculation of the adaptive sliding window, application of the MFD, execution of the machine learning algorithm), the overall detection time is calculated in an empiric way on the adopted data set. The detection time for each step also provides an indication of the related computational overhead because the computing platform used to execute the steps is the same. The estimates for all the compared approaches (i.e., entropy measures and MFD) are provided in Section 4.

4. Results

This section provides the results on the application of the proposed approach based on MFD to the CICIDS2017 data set. The results are shown in terms of the metrics described in Section 3.6. In particular, the accuracy shows the performance of the proposed approach to correctly identify the sliding windows where the attack was present or absent (i.e., legitimate traffic). Like most of the IDS data sets, the CICIDS2017 data set is heavily unbalanced because the number of legitimate windows of traffic statistics is much larger than the windows of traffic statistics where the attack was implemented. Then, the accuracy metric can be misleading. For example, a high value of accuracy may simply mean that the algorithm can correctly identify the legitimate traffic (which is the large majority) but it may fail to detect the attack. Then, the TPR and FPR are also used to evaluate the performance of the proposed approach and compare it with similar results from approaches defined in the research literature. In this particular context, the TPR indicates how well the proposed approach is able to identify the legitimate traffic in relation to the number of FN (i.e., windows of traffic predicted as attack related traffic but actually legitimate traffic). Then, if the TPR value is high, the approach suggests that there are no attacks in the analyzed traffic. Then, one approach is better than another if TPR is higher on the same data set and conditions. A higher value of precision also indicates that one approach is better to another but contrary to TPR, precision is based on the TP and the FP (in this case windows of traffic predicted as legitimate traffic but actually attack-related traffic). Instead, the FPR indicates the proportion of windows of traffic associated with attacks, which are incorrectly identified as legitimate traffic. In the IDS context, it is important that the algorithm achieves a low FPR (i.e., a lower value of FPR indicates a better performance of the approach) to minimize the risk that a window a traffic related to an attack is actually interpreted as legitimate traffic.

Finally, the detection time of an attack is shown as a relative value to the optimal approach (i.e., the approach which has the lowest computing time). The detection time is calculated as the sum of the computing steps (e.g., feature extraction) needed to detect an attack in one or more windows of traffic.

This result section is split into different sub-sections: the first Section 4.1 presents the analysis of the impact of the different hyper-parameters on the overall performance of the proposed approach. The second Section 4.2 provides the results and related analysis for the comparison of the proposed approach with similar approaches (e.g., sliding window based with entropy measures) from the literature that are not necessarily using the same data set CICIDS2017 adopted in this paper. The third Section 4.3 compares the proposed approach with the results achieved by other studies in the literature on the same CICIDS2017 adopted in this paper, but which adopt widely different approaches (e.g., machine learning or deep learning). Finally, the detection time is analyzed in Section 4.4.

As previously described in Section 3, the portion of the CICIDS2017 data set of July 03, 2017, which includes normal traffic, was used for the phase 1 of the methodology (e.g., definition of the thresholds related to legitimate traffic), while the rest of the data set was used for the phase 2 of the methodology. This is a similar approach adopted by researchers using the CICIDS2017 data set [1,29].

4.1. Optimization

This sub-section provides the results on the optimization of the hyper-parameters described in the previous sections. As discussed before, the approach is based on a number of hyper-parameters, which are summarized in Table 3.

A grid approach was used to determine the optimum values of the hyper-parameters in an 8 dimension space. While, other methods (e.g., gradient, cuckoo search algorithm) could be more efficient, it should be considered that the ranges of values for each hyper-parameter are quite limited (in most cases, they are less or equal to 6 elements). In addition, the intention is to show, in an explicit way, the impact of each hyper-parameter for the detection performance. The accuracy metric is used to determine the optimal values of the hyper-parameters.

In the following paragraphs, Tables and Figures of this Results section, it is shown how the specific detection performance is impacted by one hyper-parameter or a set of hyper-parameters while keeping the other values of the hyper-parameters constant (these are the values identified in the last column of Table 3), which are the results of the grid optimization process. The optimization process is implemented using the accuracy metric. The first optimization result is related to the adaptive sliding window algorithm and the choice of the measure and the hyper-parameters identified in Table 2. The results of the optimization are shown in Fig. 2 where it can be seen that the optimal results are achieved with the quantile measure and p=0.6. The x axis shows each of the statistical measures used in the study. The legend shows the resulting performance for each of the measures identified in Table 2 and for the different values of the hyper-parameters of the quantile measure.

As discussed before, another parameter is the time requested to execute the measure. The processing times can be extrapolated from the definition of each of the measures (i.e., computing complexity) but an empirical recording of the processing time has shown that there is not a significant difference in the computing time of each measure. If the processing time of the variance measure is used as a reference with value 1, the processing time of the quantile measure was 1.044, the burstness measure 1.045, and the entropy measure 1.736. Then, the quantile measure (which provides the optimal accuracy) has a good timing performance as well. The adaptive window operates on a range of window sizes W_S from 600 network flows to 1600 networks flow in steps of 100 network flows [600,700, …,1500,1600]. As described before, this range was chosen because of various constraints: (a) to have a lower limit of W_S to produce meaningful results for the application of MFD, (b) to have an optimal accuracy and (c) to limit the computing time needed to determine the optimal value of W_S (the larger is the range of W_S values and the more computing time is needed to identify the optimal values of W_S). The constraint (a) imposes the lower limit of $W_S = 600$, the constraints (b) and (c) are determined by evaluating different ranges of values, with the results presented in Table 4.

Fig. 2 presents the results for the optimization of the MFD hyper-parameters: g, c and k_{max}, where k_{max} is calculated with the parameter S_F and optimal W_S. The results for $g = 1$ (segment) are presented in Fig. 2(a) (note that for g=1 the c parameter does not have an influence on the calculation, but the graph has been left unaltered for consistency with the other figures), the results for $g = 2$ (square) are presented in Fig. 2(b) while the results for $g = 3$ (rhombus) are presented in Fig. 2(c). The results show that the optimal performance is obtained for $g = 3$ and $S_F = 1/160$. For all the three values of g, the detection performance decreases with the increase of the value of S_F. A potential reason for this result is that smaller values of S_F (and k_{max}) allow the MFD to capture the variations in the network flows in a more precise way (since the maximum scale is smaller) thus allowing to detect the DDoS attack with greater accuracy. The relevance of the hyper-parameter c is not so significant as the other two hyper-parameters but the value of $c = 0.0005$ has a slightly better detection performance than the other values.

An example of the evaluation of the impact of the parameter Thr_{Dec} on the detection accuracy is shown in Fig. 3 where it can be seen that the accuracy increases with larger values of Thr_{Dec} but eventually flattens for high values of Thr_{Dec} (the maximum value of the accuracy is obtained for Thr_{Dec}=3.8 for the MFD feature). This behavior is consistent with findings from other IDS studies using the sliding window approach [18,30] and it is due to the consideration that small values of Thr_{Dec} generate a large percentage of false negatives (samples

Table 3
Summary of the hyper-parameters in the proposed approach and related optimal values.

Hyper-parameter	Description	Range	Optimal value
F_A	Function used to determine the optimum sliding window size	The 4 functions described in Table 2	quantile with p=0.6
Trh_{Dec}	Threshold for detection: multiplier of the standard deviation	Range 1..4 in 0.1 steps	$Trh_{Dec} = 3.7$
N for $W_{S1}..W_{SN}$	Range of window sizes W_S	4 ranges of N: $(600 \Leftrightarrow 1200)$ $(600 \Leftrightarrow 1600)$ $(600 \Leftrightarrow 2000)$ $(600 \Leftrightarrow 2400)$	$(600 \Leftrightarrow 1600)$
D_{dec}	Distance metric used for detection	choice among Manhattan, Euclidean, Chebyschev, Minkowski p=3,4,5;	Chebyschev
HF	High level features	11 High level features defined in Table 1	MFD
MFD g	function structuring element g of MFD [5]	3 values g=1 (segment), g=2 (square), g=3 (rhombus)	g=3
MFD h	h parameter from [5]	6 values (0.05,0.01,0.005, 0.001,0.0005,0.0001)	h=0.0005
S_F (to calculate MFD k_{max})	k_{max} from [5] calculated with S_F	6 values $1, 1/20, 1/40,$ $1/80, 1/160, 1/320$	$k_{max} = 1/160$

Table 4
Accuracy of the adaptive window algorithm for different W_S ranges (W_S in 100 steps).

Measure	W_S $(600 \Leftrightarrow 1200)$	W_S $(600 \Leftrightarrow 1600)$	W_S $(600 \Leftrightarrow 2000)$	W_S $(600 \Leftrightarrow 2400)$
Entropy	0.9869	0.9893	0.9915	0.9893
Variance	0.9850	0.9863	0.9906	0.9852
Burstness	0.9777	0.9789	0.9770	09838
Quantile p=0.5	0.9877	0.9898	0.9892	0.9871
Quantile p=0.6	0.9859	**0.9930**	0.9892	0.9871
Quantile p=0.7	0.9883	0.9892	**0.9930**	0.9873
Quantile p=0.8	0.9889	0.9911	0.9909	0.9876
Quantile p=0.9	0.9887	0.9923	0.9907	0.9907

Table 5
Accuracy of the adaptive window algorithm for different distance metrics.

Distance metric D_{dec}	Accuracy
Manhattan	0.2640
Euclidean	0.8884
Minkowski $o = 3$	0.9790
Minkowski $o = 4$	0.9885
Minkowski $o = 5$	0.9917
Chebyshev	**0.9930**

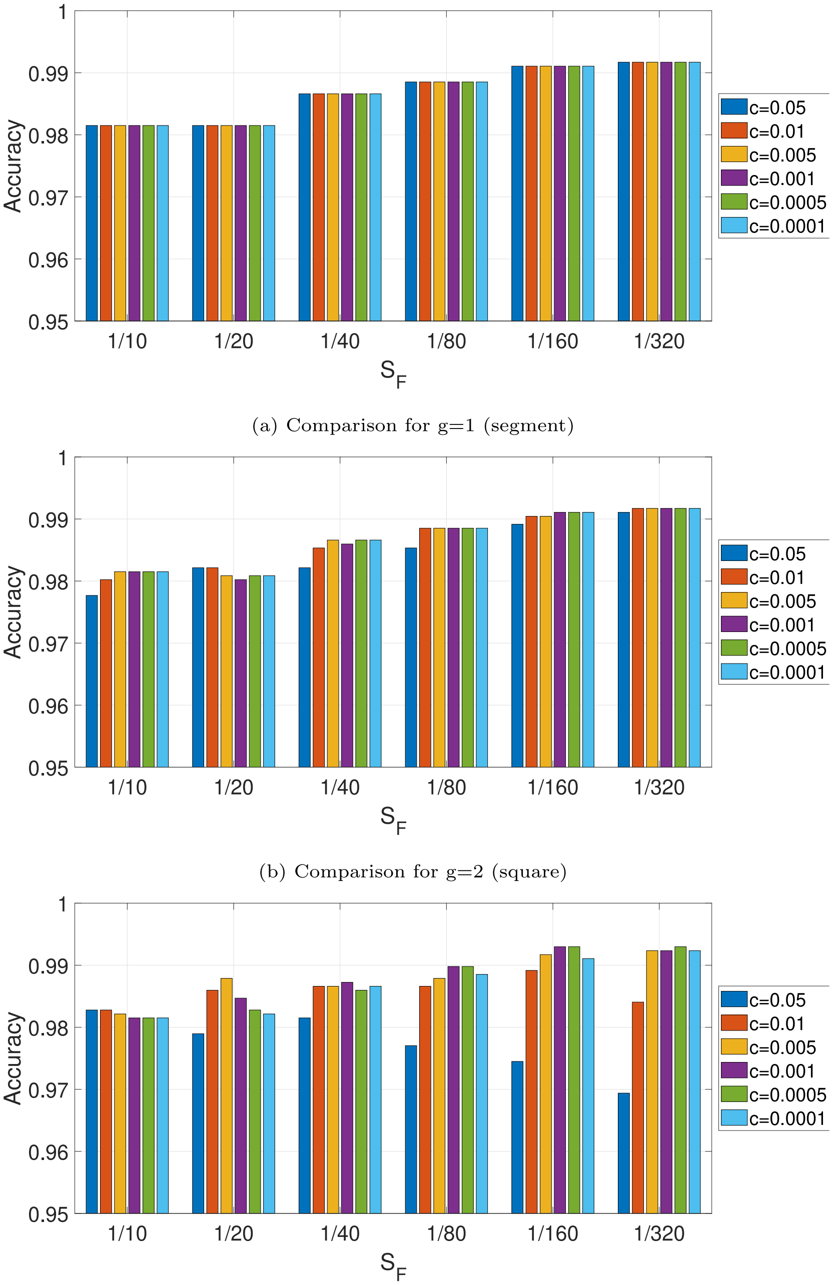
predicted to be malicious traffic but which are legitimate traffic). With increasing values of Thr_{Dec} the false negatives decrease but the probability of false positives may increase (samples predicted to be legitimate traffic but they are actually malicious traffic) to the point that the two trends balance them out and eventually make the accuracy curve flatten out.

Finally, the impact of the distance metric is evaluated and the results are presented in the following Table 5. The table shows that the Chebyshev distance manages to achieve the best classification performance while the Manhattan distance provides a very limited accuracy. We remind that the Manhattan distance, Euclidean distance and Chebyshev distance are specific cases of Minkowski distance of order $o = 1$, $o = 2$ and $o = \infty$ respectively.

4.2. Comparison among features and approaches

This sub-section compares the detection performance obtained with the MFD to the entropy or fractal measures used in other studies. Using the optimal values ($g = 3$, $c = 0.005$, Chebyshev distance and $S_F = 1/4$) obtained in the previous Section 4.1, a comparison was performed of the approach based on MFD presented in this study with the other approaches proposed in the literature: Shannon entropy in [1,4] and Renvy entropy, and Tsallis entropy in [4]. The results of the comparison for different values of W_S are presented in Fig. 4. Three values of the order of the Renvy entropy and Tsallis entropy ($r = 2, 3, 4$) are used in the comparison in a similar way to what is presented in [4]. The results shown in Fig. 4 demonstrate the superior performance of MFD in comparison to the approaches based on Shannon entropy, Renvy entropy and Tsallis entropy in a consistent way for the different values of W_S. In Fig. 4, the lines indicate the detection accuracy with the application of the adaptive method for the size W_S of the sliding window while the bar graph shows the detection accuracy obtained for each values of W_S (without using the adaptive method). Only the lines for Shannon Entropy and MFD are shown in Fig. 4 to avoid cluttering the figure. The detailed results are anyway presented in Table 6. In particular, we note that the other fractal dimension features do not provide a significant detection accuracy on this data set while the entropy measures are able to provide a relatively good performance even if it is inferior to the one based on MFD. In general, the adaptive sliding method is superior in detection accuracy in comparison to the approach with fixed value of W_S with the additional advantage that the choice of the proper value of W_S must not be determined (with additional computing cost for the identification of the optimal value of the W_S parameter).

Table 6 reports the comparison of the approach proposed in this paper and based on MFD with other high level features used in other studies. Because most of the studies are not based on the same data set used in this paper, the table does not report the metrics values of the detection performance from the other studies but it shows the detection performance obtained with the features used in the other studies for this specific data set. The results from Table 6 show that



(a) Comparison for g=1 (segment)

(b) Comparison for g=2 (square)

(c) Comparison for g=3 (rhombus)

Fig. 2. Optimization of the MFD.

the approach based on MFD can outperform significantly the other approaches presented in literature. In particular, the approach based on MFD is significantly better than the other fractal based approaches with a significant difference in accuracy (more than 20% in accuracy for the Hausdorff fractal dimension). The approaches based on entropy measures have a better performance than the other fractal dimension based approaches. In particular, the Shannon entropy and the Renyi entropy achieve a relative good detection accuracy. This may explain the fact that entropy measures are quite popular in IDS literature. In fact, the approach based on Renyi entropy p=4 can achieve even a slightly better value than the MFD based approach for the FPR (0.04 against 0.056) and Precision (0.9964 against 0.9952), but its TPR and Accuracy are worst, which may generate the risk of generating an high number of false alarms (e.g., the IDS notifies an attack but it is actually legitimate traffic) which may negatively impact the operational burden of the network administrator (i.e., he/she has to investigate the reason for the alarm, even it is a false alarm). While a difference of 2% may not seem significant, it has to be taken in consideration that network traffic in current networks amounts to millions of packets every day, which translates to ten of thousands of sliding windows in the approach proposed in this paper. Then, even an accuracy improvement of 2% (roughly equivalent to few hundreds of windows) is important from an operational point of view because it decreases considerably the number of false alarms.

4.3. Comparison with the state of art on the CICIDS2017 data set

Table 7 instead reports the comparison of the approach proposed in this paper with the results obtained by other studies on the same CICIDS2017 data set. In this case, the values of the detection metrics

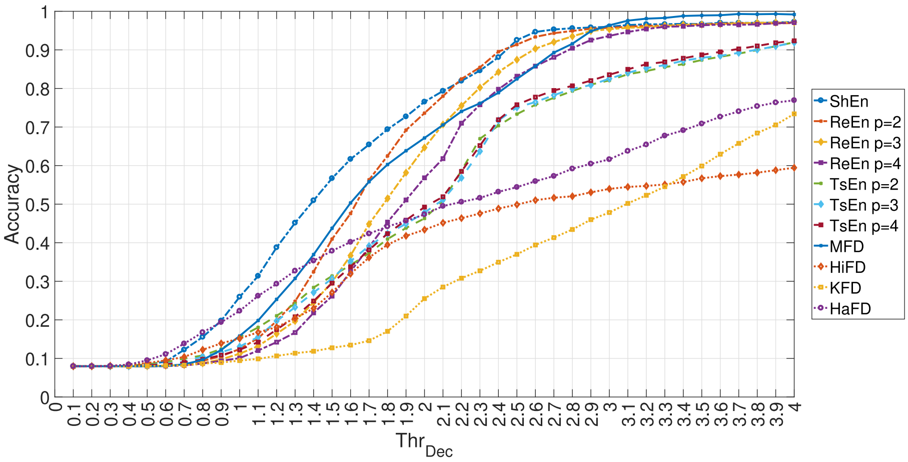


Fig. 3. Accuracy in relation to Thr_{Dec} for the eleven (11) High Level features considered in the study.

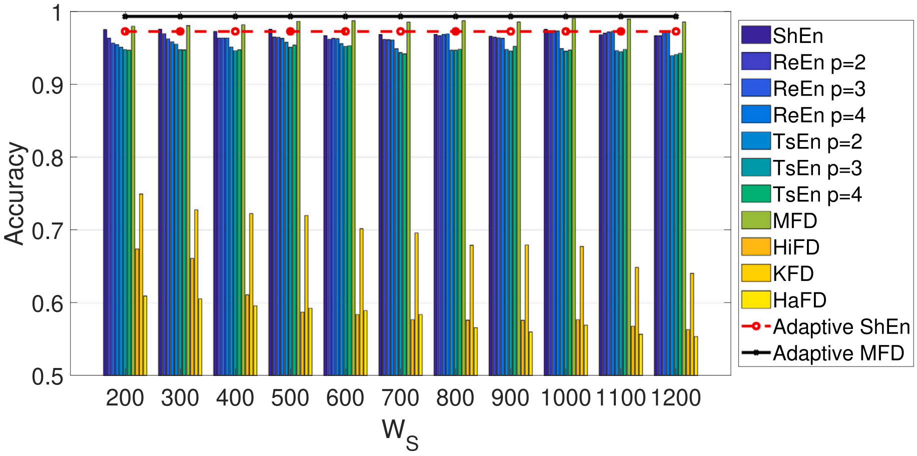


Fig. 4. Accuracy for the DDoS attack using MFD for different values of W_S in comparison to approaches based on Shannon entropy, Renvy entropy, Tsallis entropy, Higuchi fractal dimension, Katz fractal dimension and Hausdorff fractal dimension for different values of the sliding window. The straight lines on the top of the figure represent the adaptive method proposed in this manuscript for Shannon Entropy and the morphological fractal dimension.

Table 6
Comparison of the results with other approaches for IDS in literature.

Method	Accuracy	TPR recall	FPR fall-out	Precision	Ref.
Proposed method	0.9930	0.9965	0.056	0.9952	This paper.
Sliding window with Shannon entropy	0.9700	0.9889	0.2480	0.9787	[1], [4], [9],[10]
Sliding window with Renyi entropy p=2	0.9688	0.9751	0.1040	0.9908	[4], [12],[13], [20]
Sliding window with Renyi entropy p=3	0.9700	0.9737	0.0720	0.9936	[4], [12],[13], [20]
Sliding window with Renyi entropy p=4	0.9662	0.9667	0.0400	0.9964	[4], [12], [13], [20]
Sliding window with Tsallis entropy p=2	0.8992	0.9051	0.1680	0.9842	[4],[12],[14]
Sliding window with Tsallis entropy p=3	0.9005	0.9030	0.1280	0.9879	[4],[12],[14]
Sliding window with Tsallis entropy p=4	0.9101	0.9099	0.0880	0.9917	[4],[12],[14]
Sliding window with Higuchi fractal dimension	0.5816	0.5454	0	1	[15],[16]
Sliding window with Katz fractal dimension	0.6843	0.6570	0	1	
Sliding window with Hausdorff fractal dimension	0.7538	0.7471	0.1680	0.9809	[17]

Table 7
Comparison of the results with other approaches on the same CICIDS2017 data set.

Method	Accuracy	TPR recall	FPR fall-out	Precision	Ref.
Proposed method	0.9930	0.9965	0.056	0.9952	This paper.
KOAD and Shannon Entropy	0.9955	0.9524	N/A	0.9524	[1]
ML (ID3-Decision Tree)	N/A	0.98	0.2640	N/A	[29]
k-NN, RF, SVM	0.9666 (RF)	0.917	N/A	0.88	[31]
One-class SVM, Auto-encoder	0.98 (RF)	0.97	N/A	0.99	[32]

Table 8
Detection time.

Detection step	Computing time in milliseconds (ms)
Adaptive sliding window identification (variance)	34 ms
Adaptive sliding window identification (quantile)	49 ms
Adaptive sliding window identification (Shannon entropy)	59 ms
Adaptive sliding window identification (burstiness)	49 ms
High level Feature (HF) extraction (Shannon entropy) with $W_S = 600$	0.2 ms
High level Feature (HF) extraction (MFD) with $W_S = 600$	8 ms
Comparison with threshold	6 ms
Total (Shannon Entropy)	40.2 ms
Total (MFD)	48 ms

Table 9
Comparison of the computing time with machine learning approaches.

Method	Computing time	Accuracy
Proposed method (adaptive sliding window) with Shannon Entropy	1.056	0.9726
Proposed method (adaptive sliding window) with MFD	1	0.9930
Approach based on a range of fixed windows (W_S from 200 to 1600) with Shannon Entropy	2.544	0.9707 (mean value) 0.9757 (max value)
Approach based on a range of fixed windows (W_S from 200 to 1600) with MFD	2.32	0.9853 (mean value) 0.9891 (max value)
SVM with Shannon Entropy (adaptive sliding window)	1.39	0.9734
SVM with MFD (adaptive sliding window)	1.32	0.9782
RF with Shannon Entropy (adaptive sliding window)	1.08	0.9797
RF with MFD (adaptive sliding window)	1.15	0.9891

are reported directly from the published studies. We note that the proposed approach can outperform most of the results from the other studies, with the exception of the KOAD and Shannon Entropy used in [1], which has a higher accuracy. While these results may seem quite promising, it should be considered that the approach proposed in this paper uses a sliding window approach while other studies perform the classification directly on the single traffic statistics (i.e., the traffic statistics are not groups in windows). Then, the results presented in Table 7 may not be directly comparable for this reason.

4.4. Time analysis

The computing efficiency is another important parameter to evaluate the performance of the proposed approach. This section provides the computing times for the detection of the attack on the basis of the specific operations identified and described in Section 3. To summarize, the three main operations in the detection phase are:

- The selection of the optimal window using the adaptive method.
- The application of the High level Feature (HF) on the window of network traffic statistics.
- The comparison of the value of the HF with the threshold Thr_{Dec}.

The times have been calculated using a laptop with Intel core $i7 - 8550U$ working at 1.8 GHz with 16 Gigabyte of RAM. At the time of writing this paper, this computing platform is an average consumer market computer. It can be assumed that a network manager can have more powerful computing capabilities to implement IDS. To provide the time estimates, the range of sliding windows is from 200 to 1200 in 100 steps (i.e., 200,300, …,1100,1200).

The execution times are shown in Table 8, which shows that the application of the MFD to the network traffic statistics in the sliding window requires a significant amount of time in comparison to the application of Shannon entropy. On the other side, the application of the HF (either MFD or Shannon entropy or other HFs described in this study) requires less time than the execution of the adaptive sliding window algorithm. While the amount of time requested by the execution of the adaptive sliding window algorithm may seem significant, it has to be taken in consideration that it avoids the need to estimate the optimal value of W_S, which would require the execution of the entire IDS process including the initial training phase, which may take considerable time in the case of large data sets, which are typically present in the IDS context.

Then, we have also estimated the comparison between the proposed approach based on a sliding window and the application of machine learning algorithms and they are presented in Table 9. Note that the values presented in Table 9 do not take in consideration the steps described in the previous Table 8, but they provide the estimate after the feature extraction step (where the feature extraction of MFD would take considerable more time than the feature extraction with Shannon entropy) on the entire CICIDS2017 data set. Instead of presenting the absolute values of the computing time, which would be relative to the specific computing platform used in the study, we have set to 1 the minimum reported time (i.e., this approach based on MFD) and the computing times from the other methods are multiples of the minimum reported time.

Table 9 shows that the approach presented in this paper based on the MFD is the most time efficient in addition to providing the optimal detection accuracy. In particular, the machine learning methods require slightly more time to execute. Table 9 shows also the significant advantage of the adaptive window approach as the tuning of the sliding window size may require more than twice the time of the adaptive approach. Note that the machine learning approach based on the sliding window uses the samples for classification where the high level features HF are calculated on the adapted windows.

5. Conclusions

This paper has investigated the application of Morphological Fractal dimension (MFD) in combination with a sliding window approach to the problem of detection of Distributed Denial of Service (DDoS) attack in fixed networks. Network traffic statistics are collected in windows on which the MFD is applied. The underlying concept is that attacks modify the structure of the network traffic, which in turn will change the value of the MFD. This approach has been validated in literature for other types of features like entropy measures (e.g., Shannon entropy, because the entropy of the network traffic is also changed in presence of a DDoS attack), but it is the first time that MFD is applied to this problem to the best knowledge of the authors. Because a fixed length of the sliding window may not capture well the duration and characteristics of the DDoS attack, this paper combines the MFD with an adaptive sliding window method, where the optimal size of the window is estimated in an adaptive way. The proposed approach is applied to the recent intrusion detection data set CICIDS2017 and compared with other features derived from literature like the Shannon entropy, Renyi entropy, and Tsallis entropy or other fractal dimension measures. The results show that the approach based on MFD outperforms the other approaches at the cost of a higher computing time to extract the MFD value from each sliding window. On the other side, the detection time is still relatively short (i.e., 48 ms) even on a consumer market laptop using the public CICIDS2017 data set. We have also evaluated the performance of the application of MFD to the problem of the detection of Distributed Denial of Service (DDoS) attacks for different values of the hyper-parameters of the definition of MFD.

This paper was focused on the analysis of the detection performance of MFD using a simple sliding window approach without the attempt to combine it with more sophisticated anomaly detection methods or machine learning algorithms (e.g., deep learning), which may be able to obtain even a higher detection performance (as it was done in [1] with Shannon entropy). This will be the topic of future developments.

CRediT authorship contribution statement

Gianmarco Baldini: Conceptualization, Software, Methodology, Writing – original draft, Writing – review & editing. **Irene Amerini:** Methodology, Writing – original draft, Writing – review & editing.

Declaration of competing interest

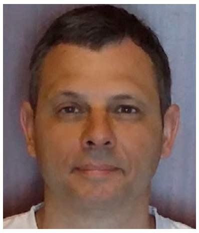
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding source

This work has been partially supported by the European Commission through project SerIoT funded by the European Union H2020 Programme under Grant Agreement No. 780139. The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the European Commission.

References

[1] S.D. Çakmakçı, T. Kemmerich, T. Ahmed, N. Baykal, Online DDoS attack detection using Mahalanobis distance and kernel-based learning algorithm, J. Netw. Comput. Appl. 168 (2020) 102756.
[2] H.-J. Liao, C.-H.R. Lin, Y.-C. Lin, K.-Y. Tung, Intrusion detection system: A comprehensive review, J. Netw. Comput. Appl. 36 (1) (2013) 16–24.
[3] T.F. Lunt, A survey of intrusion detection techniques, Comput. Secur. 12 (4) (1993) 405–418.
[4] P. Bereziński, B. Jasiul, M. Szpyrka, An entropy-based network anomaly detection method, Entropy 17 (4) (2015) 2367–2408.
[5] P. Maragos, F.-K. Sun, Measuring the fractal dimension of signals: morphological covers and iterative optimization, IEEE Trans. Signal Process. 41 (1) (1993) 108.
[6] M. Ring, S. Wunderlich, D. Scheuring, D. Landes, A. Hotho, A survey of network-based intrusion detection data sets, Comput. Secur. 86 (2019) 147–167.
[7] N. Moustafa, J. Hu, J. Slay, A holistic review of network anomaly detection systems: A comprehensive survey, J. Netw. Comput. Appl. 128 (2019) 33–55.
[8] B.B. Zarpelão, R.S. Miani, C.T. Kawakani, S.C. de Alvarenga, A survey of intrusion detection in Internet of Things, J. Netw. Comput. Appl. 84 (2017) 25–37.
[9] S. Behal, K. Kumar, M. Sachdeva, D-FACE: An anomaly based distributed approach for early detection of ddos attacks and flash events, J. Netw. Comput. Appl. 111 (2018) 49–63.
[10] J.-H. Jun, C.-W. Ahn, S.-H. Kim, DDoS attack detection by using packet sampling and flow features, in: Proceedings of the 29th Annual ACM Symposium on Applied Computing, 2014, pp. 711–712.
[11] J. David, C. Thomas, DDoS attack detection using fast entropy approach on flow-based network traffic, Procedia Comput. Sci. 50 (4) (2015) 30–36.
[12] C. Callegari, S. Giordano, M. Pagano, An information-theoretic method for the detection of anomalies in network traffic, Comput. Secur. 70 (2017) 351–365.
[13] M.H. Bhuyan, D.K. Bhattacharyya, J.K. Kalita, Information metrics for low-rate DDoS attack detection: A comparative evaluation, in: 2014 Seventh International Conference on Contemporary Computing, IC3, IEEE, 2014, pp. 80–84.
[14] X. Ma, Y. Chen, DDoS detection method based on chaos analysis of network traffic entropy, IEEE Commun. Lett. 18 (1) (2013) 114–117.
[15] Z. Xia, S. Lu, J. Li, DDoS flood attack detection based on fractal parameters, in: 2012 8th International Conference on Wireless Communications, Networking and Mobile Computing, IEEE, 2012, pp. 1–5.
[16] V. Bulavas, Fractal dimensionality of network traffic as a feature for intrusion detection, in: 11th International Workshop on Data Analysis Methods for Software Systems, DAMSS 2019, Druskininkai, Lithuania, November 28-30, 2019, Vilnius University Press, 2019.
[17] Y. Labit, J. Mazel, Hidden: Hausdorff distance based intrusion detection approach dedicated to networks, in: 2008 the Third International Conference on Internet Monitoring and Protection, IEEE, 2008, pp. 11–16.
[18] G. Baldini, On the application of entropy measures with sliding window for intrusion detection in automotive in-vehicle networks, Entropy 22 (9) (2020) 1044.
[19] B.D. Fulcher, N.S. Jones, Highly comparative feature-based time-series classification, IEEE Trans. Knowl. Data Eng. 26 (12) (2014) 3026–3037.
[20] R. Yan, G. Xu, X. Qin, Detect and identify DoS attacks from flash crowd based on self-similarity and Renyi entropy, in: 2017 Chinese Automation Congress, CAC, IEEE, 2017, pp. 7188–7194.
[21] J. Monge-Álvarez, Maragos-Sun fractal dimension (2021), URL https://www.mathworks.com/matlabcentral/fileexchange/51175-margaos-sun-fractal-dimension.
[22] J. Monge-Álvarez, Higuchi and Katz fractal dimension measures (2021), URL https://www.mathworks.com/matlabcentral/fileexchange/50290-higuchi-and-katz-fractal-dimension-measures.
[23] A.F. Costa, G. Humpire-Mamani, A.J.M. Traina, An efficient algorithm for fractal analysis of textures, in: 2012 25th SIBGRAPI Conference on Graphics, Patterns and Images, IEEE, 2012, pp. 39–46.
[24] B.B. Mandelbrot, B.B. Mandelbrot, The Fractal Geometry of Nature, Vol. 1, WH freeman New York, 1982.
[25] P. Frankhauser, The fractal approach. a new tool for the spatial analysis of urban agglomerations, Popul. Engl. Sel. (1998) 205–240.
[26] B.B. Chaudhuri, N. Sarkar, Texture segmentation using fractal dimension, IEEE Trans. Pattern Anal. Mach. Intell. 17 (1) (1995) 72–77.
[27] W. Iqbal, J.L. Berral, D. Carrera, et al., Adaptive sliding windows for improved estimation of data center resource utilization, Future Gener. Comput. Syst. 104 (2020) 212–224.
[28] K.-I. Goh, A.-L. Barabási, Burstiness and memory in complex systems, Europhys. Lett. 81 (4) (2008) 48002.
[29] I. Sharafaldin, A.H. Lashkari, A.A. Ghorbani, Toward generating a new intrusion detection dataset and intrusion traffic characterization, in: ICISSP, 2018, pp. 108–116.
[30] S. Ohira, A.K. Desta, I. Arai, H. Inoue, K. Fujikawa, Normal and malicious sliding windows similarity analysis method for fast and accurate IDS against DoS attacks on in-vehicle networks, IEEE Access 8 (2020) 42422–42435.
[31] M. Aamir, S.M.A. Zaidi, Clustering based semi-supervised machine learning for DDoS attack classification, J. King Saud Univ.-Comput. Inf. Sci. (2019).
[32] H. Attak, M. Combalia, G. Gardikis, B. Gastón, L. Jacquin, D. Katsianis, A. Litke, N. Papadakis, D. Papadopoulos, A. Pastor, et al., Application of distributed computing and machine learning technologies to cybersecurity, Space 2 (2018) I2CAT.



Gianmarco Baldini obtained the Laurea degree in electronic engineering at the University of Rome, Italy in 1993 and the Ph.D. degree in Computer Science in 2019 at the University of Insubria, Italy. He has worked in the Research and Development departments of large multi-national companies in the field of wireless communications and ICT in Italy, UK, Ireland and USA before joining the Joint Research Centre (JRC) of the European Commission in 2007. In the JRC, he has been working in various areas including wireless communications, security, positioning and machine learning and he contributed to the formulation of European policies in the areas of radio frequency spectrum, road transportation and cybersecurity. He has co-authored more than 70 peer-reviewed papers in international journals and conferences.

Irene Amerini received the Laurea degree in computer engineering and the Ph.D. degree in computer engineering, multimedia, and telecommunication from the University of Florence, Italy, in 2006 and 2010. She is currently Assistant Professor at Sapienza University of Rome, Italy. She has received the Italian Habilitation for Associate Professor in telecommunications and computer science. She was a Visiting Scholar with Binghamton University, NY, USA, in 2010. She has been a Visiting Research Fellow of the School of Computing and Mathematics, Charles Sturt University, Australia, since 2018, offered by the Australian Government Department of Education and Training, through the Endeavour Scholarship & Fellowship Program. Her main research activities include digital image processing, multimedia content security technologies, secure media, and multimedia forensics. She is a member of the IEEE Information Forensics and Security Technical Committee and EURASIP TAC Biometrics, Data Forensics, and Security. She is an Associate Editor of the IEEE ACCESS and a Guest Editor of several international journals.