Detecting Denial-of-Service Attacks Using the Wavelet Transform

Mohamed Hamdi, Noureddine Boudriga

Communication Networks and Security Research Lab.
University of Carthage, Tunisia

Abstract

Anomaly-based intrusion detection is a crucial research issue as it permits to identify attacks that does not necessarily have known signatures. However, approaches using anomalies often consume more resources than those based on misuse detection and have a higher false alarm rate. This paper presents an efficient anomaly analysis method that is proved to be more efficient and less complex than the existing techniques. The approach relies on monitoring the security state by using a set of accurate metrics. The Wavelet Transform (WT) is used to decompose these metrics in the time-scale space. Attacks are viewed as Lipschitz singularities that arise in some specific points of time. Henceforth, the anomaly detection process is performed through processing the signals representing the metrics. The proposed approach is also shown to be extensible to the case where the monitoring points, used to gather the measurable features, are distributed according to the network topology.

Keywords Anomaly detection, wavelet transform, Lipschitz regularity, intrusion detection, Denial-of-Service attacks.

1. Introduction

Over the past few years, the rapid evolution of network connectivity and service accessibility technologies has favored the increase of intrusions and attacks. Due to the amount of losses that have resulted from the experienced network attacks, communication systems' security has been a growing concern. Especially, the problem of detecting intrusions, attacks and other forms of network abuses and misuses is of prominent importance. This is because one can assume that the efficiency of the incident response system mainly relies on the accuracy of the associated detection techniques. Roughly speaking, intrusions could be considered as transitions between secure and non-secure system states. However, in practice, characterizing such transitions turns out to be a complex task. The available intrusion detection approaches can be divided in two categories: pattern-based detection and anomaly-based detection [1]. While the former methods, also referred to as misuse detection, locate security problems by examining patterns of users activities within flows, logs, and usage files, the latter class looks for deviation in normal usage behavior of the network, information systems, and user profiles. Even though some of these methods are witnessing an increasing interest, no method can catch all types of intrusions.

Denial-of-Service (DoS) attacks represent a particular interest from the intrusion detection perspective. Recent statistical surveys [2,3] show that DoS ranks at the fourth place in the list of the most virulent attack classes against information systems. In [4], more than 90% of the respondent Internet Service Providers (ISPs) reported that DoS flooding attacks "dominated all observed attacks". This means that DoS attacks constitute the most significant threats against the network infrastructure. From a technical point of view, DoS attacks exhibit many important features:

(i) DoS attacks can involve thousands of computers. An attacker can compromise a set of intermediate hosts (zombies) that will launch the attack process. Obviously, increasing the number of zombies exacerbate the efficiency of the DoS attack and early detection is therefore harder,

(ii) Unlike other attacks, the network flow used to carry out a DoS attack slightly differs from a normal traffic. When analyzing packet headers and payloads, the Intrusion Detection System (IDS) can hardly realize that an intrusion is in progress. Hence, anomaly-based detection strategies are more suitable to detect DoS attacks.

(iii) The detection system can be bypassed by the attacker when the attack flow can not be processed by the IDS. This means that the DoS detection approach should be possible to be implemented at a reasonable numerical cost.

This paper describes the design of an anomaly-classifier system for DoS detection and addresses the related mathematical issues. The classifier can monitor the activities of the computer network at multiple levels (from traffic engineering to user activity levels) and determines the correlation among the monitored parameters (or metrics). The major role of this classifier is to identify attack-related anomalies. It basically allows to differentiate between 'virulent' and 'benign' anomalies. Our method is based on the concept of period of observation and uses wavelet theory. It has three major advantages with respect to the existing methods: first it does not require the storage of users profiles, data files on attacks, or statistical usage. Second, it offers a reduced complexity of the form $O(n)$, where n is the size of the observation periods. Last, unlike Bayesian techniques, it does not require an a priori information about the monitored anomalies.

The remaining part of this paper is organized as follows: Section 2 describes the commonly used techniques to represent and detect computer network anomalies. Section 3 gives a formal representation of these anomalies. Section 4 explains how anomaly detection can be performed using wavelet theory and presents a case study where wavelet-based classification is demonstrated. Section 5 concludes this paper.

2. Existing DoS detection techniques

This section discusses approaches for detecting DoS attacks. We first underline the basic features of DoS attacks and select the most appropriate detection strategy with respect to the attack process. Next, we present a review of the existing statistical techniques and highlight their shortcuts.

2.1. *Denial-of-Service attacks*

A DoS attack basically attempts to cripple a service which is available online. The outcome of a DoS attack can vary from degrading the service availability to denying the service to its users (i.e., service disruption). A DoS attack can be performed using various strategies. In [5], the main classes of DoS techniques are given depending on the type of victim:

– Application attacks are directed against a specific application that runs on the victim host. This means that the remaining services associated to this host would remain available.
– Host attacks disrupt access to the victim host. For instance, this can be achieved by crashing or rebooting the machine. The attacker can also send a huge packet flow to the target machine to flood its TCP/IP stack.
– Resource attacks are conducted against sensitive points of a specific network. Typical victim nodes include DNS servers and network routers. The impact of these attacks is important with regard to the number of hosts that may be affected by the failure of the victim focal component.
– Network attacks deploy randomly generated packets in the victim network so that a high proportion of the available bandwidth is consumed. This results in a noticeable degradation in terms of Quality-of-Service (QoS).
– Infrastructure attacks aim at simultaneously harm multiple components of a distributed service. The major characteristic of these DoS attacks is the coordination of the elementary attacks that are directed against the individual components.

The objective of intrusion detection is to monitor network traffic and generate alerts when an attack occurs. Two principal detection mechanisms are often considered: signature-based detection and anomaly based-detection. Signature based-detection is closely related to pattern recognition as the sniffed traffic is compared to a set of known attacks. The efficiency of this approach depends essentially on the number of attack signatures that the system knows. In fact, it does not generate an alert for an attack that has not a corresponding signature. In the second mechanism, detection results depend on the values of several measurable features, called *metrics*. This assumes that system's normal behavior can be described. Therefore, what deviates from this proper behavior is the source of an alert. In the following, we give an overview of several anomaly-based detection techniques that have been proposed in the literature for detecting DoS attacks. Two categories of techniques are commonly used in this context [5]: *change-point detection (thresholding)*, and *activity profiling*.

2.2. *Change-point detection*

The first class of anomaly-based detection techniques is called thresholding and is based on the comparison of certain attributes of the system to values that correspond to the boundary of the normal events space. To this end, the detection engine should encompass a sensor system that gathers relevant data (with respect to security alerts). In practice, threshold detection is a small component of an IDS and is seldom used as an independent system. In fact, it is generally based on a limited set of metrics that do not express

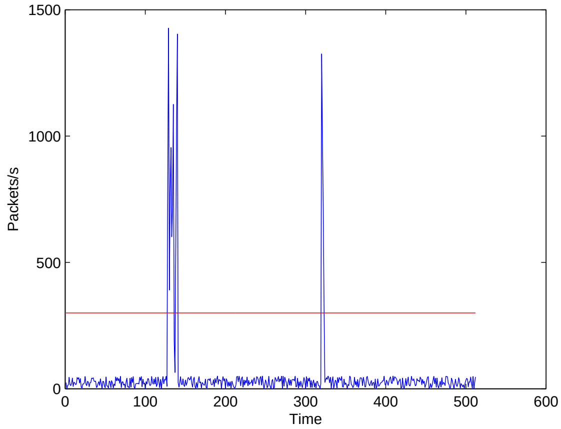

Fig. 1. Number of packets per second received by a victim server.

completely the state of the monitored network.

A question that needs to be addressed when designing a threshold-based IDS is how to set the threshold value. In fact, the detection process can return a high rate of false positives if the threshold value is too low, or a great amount of false negatives if it is excessively high. Figure 1 represents a case where intrusion detection is performed based on the number of packets received per time interval. The reader would have noticed that the resulting signal includes two peaks. The first one corresponds to an attack while the second is caused by a high demand on the server. Clearly, if the threshold value is fixed to the value depicted by the red line, both of peaks would be erroneously characterized as intrusions.

In most cases, Bayesian decision theory is used to select the optimal threshold [6,7]. According to this theory, a threshold is said optimal if it minimizes the Bayes risk function that is computed using the false positive and the false negative probabilities. The major disadvantage of this approach is that one of its fundamental hypotheses can hardly be verified in our context. In fact, to compute the Bayesian risk, an a priori knowledge about the distribution of the network attacks is needed. As the occurrence of network attacks is mainly related to human factors, it is not easy to statistically model it even by the use of historical data.

2.3. *Activity profiling*

Statistical techniques for anomaly detection have constituted a focal topic in the intrusion detection literature. They basically consist in generating a periodically updated profile of the system's normal behavior by the use of historical data (frequency tables, means, entropies, etc.). Then, during each observation period, the monitored packet flow is classified as "normal" or "abnormal".Ye *et al.* [9,10] discussed probabilistic techniques of intrusion detection, including decision trees, Hotelling T^2 test, chi-square multivariate test and Markov Chains. These tests are applied to the gathered data to investigate its frequency property and its ordering property.

Taylor *et al.* [11], [12] present a method for detecting network intrusions that addresses the problem of monitoring high speed network traffic and the time constraints on administrators for managing network security. They use multivariate [11,12] statistics techniques, namely, Cluster Analysis and Principal Component Analysis to find groups in the observed data.

DuMouchel *et al.* [13] discuss a method for detecting unauthorized users masquerading as a registered user by comparing in real time the sequence of commands given by each user to a profile of the users past behavior. They use a Principal Component Regression model to reduce the dimensionality of the test statistics.

Staniford-Chen *et al.* [14] address the problem of tracing intruders who obscure their identity by logging through a chain of multiple machines. They use Principal Component Analysis to infer the best choice of thumbprinting parameters from data. They introduce thumbprints, which are short summaries of the content of a connection.

Shah *et al.* [15] study how fuzzy data mining concepts can cooperate in synergy to perform Distributed Intrusion Detection. They describe attacks using a semantically rich language, reason over them and subsequently classify them as instances of an attack of a specific type. They use Principal Component Analysis to reduce the dimensionality of the collected data.

2.4. *Limits of the existing approaches*

The major shortcut of classic anomaly-detection techniques is that they are not built upon a strong relationship between anomalies and attacks. This can lead to a high false alarm rate as many abnormal values of a measured signal can correspond to a normal behavior of the system under analysis. For instance, consider a web server that was victim of a SYN/flooding attack (see Example 1). If the security analyst observes the number of packets per second that the server receives, he would notice a peak (abnormal value) at the moment of the attack. However, and as illustrated in Figure 1, a peak does not always imply the occurrence of a malicious action. Indeed, while the first peak corresponds to the attack mentioned above, the second peak results from a quasi-simultaneous connection of big amount of clients. The latter event is normal as it occurs at 08:00 a.m. (beginning of office hours), a time when most of employees connect to the website from their offices. The conclusion from this example is that the detection of abrupt changes or abnormal values is not sufficient to characterize computer network attacks.

Therefore, we need an other detection mechanism that should be more adapted to network security in the sense that it takes into consideration the tight difference between anomalies and attacks. This mechanism should also have a low computational complexity because of the limited resources of the intrusion detection components.

To address this issue, we propose to model attacks as a noise that affects a signal, or a set of signals, that represent the normal behavior of the system. The benefit from this reasoning is that several techniques commonly used in signal processing to detect and reduce different kinds of noise can be adapted to our context. In the frame of our work, we focused on wavelet theory and Lipschitz regularity as will be detailed in Section 4. However, to apply these methods, a convenient modeling of network anomalies is needed.

3. A mathematical framework for anomaly representation

Typically, in a networked infrastructure, it is hard to determine whether the system is in a secure state or not at a given point in time. Most of the aforementioned DoS detection mechanisms rely on several metrics (i.e., measurable information) to elucidate this point. In this section, we introduce a mathematical framework to represent anomalies based on a set of metrics.

3.1. *Metrics representation*

Change-point techniques for intrusion detection can be viewed as a decision problem where metrics value are used to characterize the system state. Supposing that a function $f(.)$ models the behavior of a specific metric, the decision process can not be conducted without translating $f(.)$ to a domain where the classification of system states is trivial. To this purpose, we introduce the concept of *detector transform* defined in the following.

Definition 1 *Let $I = A \biguplus N$ be a time interval and $f(.)$ be an observable function (metric) defined on I. A and N denote the instants corresponding the occurrence of an abnormal event and to the occurrence of no abnormal events, respectively. A mathematical transform Θ is said to be a pseudo-detector transform if, and only if:*

$$\Theta \circ f(I) = A_{\Theta \circ f} \biguplus N_{\Theta \circ f}.$$

Moreover, if $\Theta \circ f(A) = A_{\Theta \circ f}$ then Θ is called a detector transform.

According to this definition, a detector transform allows to identify the non-secure system states based on its output values. However, detector transforms are generally impossible to implement because they require a zero-error detection mechanism. Therefore, anomaly-based intrusion detection often relies on pseudo-detector transforms.

As an extension to this reasoning, several tools allowing to compare detector transform should be built.

The first task in the anomaly representation is to define the metrics that have to be monitored. Consider three families of functions denoted by $(\upsilon_i(t))_{i\in\{1,..,\Upsilon\}}$, $(\eta_j(t))_{j\in\{1,..,\mathbf{H}\}}$ and $(\nu_k(t))_{k\in\{1,..,\mathbf{N}\}}$ that model the following activities respectively:

- User-level activities. These include attributes, such as most used commands, typing frequency, and login/logout period, that help developing the profiles of behavior patterns of users.
- Host-level activities. These involve attributes (such as file structure, consumed CPU, and consumed memory) that provide indication of resource usage.
- Network-level activities. These use attributes (such as total packet count, packets with specific source or destination ports, and packets with specific source or destination addresses) to provide information that are gathered on network usage and engineering.

The above functions are the metrics that represent efficiently the information system state. They should be measured and monitored continuously. Information system administrators have to determine what attributes to record to ensure efficient representations of their systems. Moreover, for each of these metrics, a set of decision rules have to be defined in order to decide whether a value taken by a given function corresponds to a misuse or a legitimate use of the system. To this purpose, we introduce the following sets:

- $(N_{\upsilon_i})_{i\in\{1,..,\Upsilon\}}$ (resp. $(A_{\upsilon_i})_{i\in\{1,..,\Upsilon\}}$): the sets of normal (resp. abnormal) values that could be taken by the functions $(\upsilon_i(t))_{i\in\{1,..,\Upsilon\}}$,
- $\left(N_{\eta_j}\right)_{j\in\{1,..,\mathbf{H}\}}$ (resp. $\left(A_{\eta_j}\right)_{j\in\{1,..,\mathbf{H}\}}$): the sets of normal (resp. abnormal) values that could be taken by the functions $(\eta_j(t))_{j\in\{1,..,\mathbf{H}\}}$,
- $(N_{\nu_k})_{k\in\{1,..,\mathbf{N}\}}$ (resp. $(A_{\nu_k})_{k\in\{1,..,\mathbf{N}\}}$): the sets of normal (resp. abnormal) values that could be taken by the functions $(\nu_k(t))_{k\in\{1,..,\mathbf{N}\}}$.

An action at an instant t_0, denoted $\tau(t_0)$, can force the modification of the values of the aforementioned functions at t_0.

Definition 2 *An action $\tau(t_0)$ is said to be* **anomalous** *if one of the following conditions holds:*

(i) $\exists i \in \{1,..,\Upsilon\}$ *such that* $\upsilon_i(t_0) \in A_{\upsilon_i}$,
(ii) $\exists j \in \{1,..,\mathbf{H}\}$ *such that* $\eta_j(t_0) \in A_{\eta_j}$,
(iii) $\exists k \in \{1,..,\mathbf{N}\}$ *such that* $\nu_k(t_0) \in A_{\nu_k}$.

This means that if, during a action, a metric takes an abnormal value, then the whole transaction is abnormal. Therefore, the efficiency of this decision rule depends mainly on the efficiency of the elementary decision rules stating whether the value of a single metric is normal or not. The characterization of network anomalies is then easy as it consists simply in checking if the values of the monitored metrics belong, at a given instant, to several predefined sets. A more sensitive problem is to state whether an anomaly is related to an attack or not.

3.2. *The attack/singularity analogy*

In its broadest definition, a computer attack is any malicious action performed against a computer system or the services it provides. In [23], B. Schneier describes computer attacks as *exceptions* or *events that take people by surprise*. This feature makes the detection of these attacks a difficult task. In fact, the detection system is often faced with the sensitive problem of stating if an event corresponds to a malicious action or not without having enough knowledge to do.

The aforementioned definition of computer attacks is particularly convenient in our context. In fact, most of these attacks can be assimilated as abrupt changes in some measurable signals. Thus, anomaly detection can be performed through the detection of singularities in a set of significant metrics. Obviously, the choice of these metrics is of primal importance as it has a big influence on the performance of the detection system. However, this issue is out of the scope of this paper. In the following, we give two examples illustrating the selection of efficient metrics depending on the nature of the attack.

Example 1 SYN flood (Neptune) attack

This is a DoS attack to which most of TCP/IP implementations are vulnerable. It consists in opening enough half-open TCP connections to exceed the capacity of the victim machine who will be unable to accept more connections. This attack can be efficiently detected by monitoring the number of open connections on the protected machine. Singularities can be stated when the number of connections exceeds a predefined threshold.

Example 2 Mailbombing

A mailbomb is an attack in which many messages are sent to a user registered on the mail server overflowing its queue and causing the failure of his account. It is rather harmful to particular users than to the whole server. It can be detecting through the measurement of the number of messages (or the amount of bytes) received by the server per time-unit.

In these two examples, anomalies can be easily identified by detecting the exceeding of the monitored metrics to a specific value. Nonetheless, to state if an anomaly is an attack or not, a deeper mathematical analysis of the anomaly is required. A detailed study of the first example is given in Section 4.

4. Wavelets and Lipschitz regularity

Wavelets can decompose one-dimensional signals to analyze both their special frequencies and time localization. This section presents a wavelet-based tool for detecting singularities (in terms of Lipschitz regularity) of a finite energy functions. We briefly review the basic concepts related to wavelet theory (the interested reader would refer to [18] for more details). Then, we give several results, that have been proved by Mallat and Hwang in [22], highlighting the benefit of using wavelets instead of the Fourier transform to detect singularities.

4.1. *Wavelet theory fundamentals*

A function $\Psi(.)$ belonging to $\mathcal{L}^2(\mathbb{R})$ (Holder space of functions with finite energy) and centered around zero (i.e., $\int_{-\infty}^{+\infty}\Psi(t)dt = 0$) is said to be a wavelet if, and only if, its Fourier transform $\widehat{\Psi}(\omega)$ satisfies the following condition:

$$\int_{-\infty}^{+\infty}\frac{\left|\widehat{\Psi}(\omega)\right|^2}{\omega}d\omega \;<\; +\infty. \tag{1}$$

$\Psi(.)$ is also called the **"mother wavelet"**.

The wavelet transform of a function $f(.) \in \mathcal{L}^2(\mathbb{R})$ is defined by:

$$W_f(s,t) = f*\Psi_s(t) = \int_{-\infty}^{+\infty} f(u)\Psi_s(t-u)du, \tag{2}$$

where $s \in \mathbb{R}_+^*$ is the scale factor, $*$ is the convolution operator and $\Psi_s(t) = \frac{1}{\sqrt{s}}\Psi(\frac{t}{s})$ is the dilation of the wavelet $\Psi(.)$ by s.

It offers an alternative to the windowed Fourier transform (or Short-Time Fourier Transform, STFT) for non-stationary signal analysis. The principal shortcut of the STFT is that is uses a constant-length window that does not permit a spatial analysis of the signal. On the other hand, in the wavelet transform, wide windows are applied for low frequencies and short windows for high frequencies. This provides an idea about the global and the local properties of the signal $f(.)$.

Often in practice, $W_f(s,t)$ are computed for discrete values of s and t. The main constraint when choosing these values is to ensure that the transform is not redundant. According to Equation 2, this requirement can be satisfied if the functions $(\Psi_{s,u}(t))_{s\in\mathbb{R}_+^*,u\in\mathbb{R}}$ given by:

$$\forall(s,u) \in \mathbb{R}_+^*\times\mathbb{R}, \Psi_{s,u}(t) = \Psi_s(t-u) = \frac{1}{\sqrt{s}}\Psi(\frac{t-u}{s}), \tag{3}$$

are a basis of $\mathcal{L}^2(\mathbb{R})$.

The Digital Wavelet Transform (DWT) assumes the computation of wavelet coefficients for discrete scale factor $s = 2^{-j}$ and translation $u = k2^{-j}$ for $j, k \in \mathbb{Z}$. In fact, these values of the wavelet transform parameters define an orthogonal basis $\left(\Psi_{j,k}(t) = 2^{-\frac{j}{2}}\Psi(2^{-j}t - k)\right)_{j,k\in\mathbb{Z}}$ called *the wavelet basis*.

To represent a function $f(.)$ at different resolutions through the use of the "mother wavelet", we have to introduce the scaling function verifying:

$$\psi(t) = \sqrt{2}\sum_{k=-\infty}^{+\infty} g(k)\phi(2t - k), \qquad (4)$$

where $g(k) = \frac{1}{\sqrt{2}} < \psi(t), \phi(2t - k) >$.

In [19], Mallat proposed a description of a discrete orthonormal wavelet transform that is based on the concept of multiresolution analysis. He demonstrated that a function $f(.)$ can be written as:

$$f(t) = \sum_{k\in\mathbb{Z}} \lambda_{-r+1,k}\varphi_{-r+1,k}(t) + \sum_{j=-r+1}^{j=-1}\sum_{k\in\mathbb{Z}} \gamma_{j,k}\psi_{j,k}(t), \qquad (5)$$

where $\lambda_{i,j}$ and $\gamma_{i,j}$ represent the projection coefficients of f on orthogonal complementary sub-spaces of $\mathcal{L}^2(\mathbb{R})$.

This brings to evidence the main feature of the wavelet transform: the representation of a function at a given resolution (or scale) can be obtained from its representation at a coarser resolution. The transform is then said to be multiscale.

A similar development has been conducted to perform multiresolution decomposition in $l^2(\mathbb{Z})$. Basically, it consists in applying the same pyramidal decomposition to the original signal [20].

4.2. *Wavelet-based singularity detection*

In mathematics, the local regularity of functions is often evaluated by the use of Lipschitz exponents.

Definition 1 Local Lipschitz regularity, uniform Lipschitz regularity

A function $f(.)$ is said to be Lipschitz-α, for $0 \leq \alpha \leq 1$, at a point t_0, if, and only if, there exists a constant A such that for all points t in a neighborhood of t_0

$$|f(t) - f(t_0)| \leq A\,|t - t_0|^{\alpha}. \qquad (6)$$

$f(.)$ is uniformly Lipschitz-α over the interval $]a, b[$ if there exists a constant A such that Equation 6 holds for every $(t_0, t) \in\,]a, b[^2$.

Furthermore, $f(.)$ is said to be singular in t_0 if it is not Lipschitz-1 in t_0.

It is noteworthy that this definition can be extended to $\alpha > 1$.

Lipschitz regularity can be characterized through the use of the Fourier transform. In fact, a function $f(.)$ is bounded and uniformly Lipschitz-α over $\mathbb{R}$ if

$$\int_{-\infty}^{+\infty} \left|\widehat{f}(\omega)\right| (1 + |\omega|^{\alpha})\, d\omega \; < \; +\infty. \qquad (7)$$

Unfortunately, Equation 7 is not applicable to address local signal properties because it only allows to state about uniform regularity on a specific domain. In [21], Jaffard proposed a result giving necessary and sufficient conditions for characterizing local Lipschitz regularity using the wavelet transform. He demonstrated that if a wavelet $\Psi(.) \in \mathcal{L}^2(\mathbb{R})$ and a function $f(.) \in \mathcal{L}^2(\mathbb{R})$ verify:

- $\Psi(.)$ is n-times continuously differentiable,
- $\Psi(.)$ has n vanishing moments ($\int_{-\infty}^{+\infty} t^k\Psi(t)dt = 0$ for all $0 \leq k \leq n + 1$),
- $\Psi(.)$ has a compact support,
- $f(.) \in \mathcal{L}^2(\mathbb{R})$ is Lipschitz-α at a point t_0, $0 \leq \alpha \leq n$,

then there exists a constant A such that for all points t in a neighborhood of t_0 and any scale s

$$|W_f(s, t)| \leq A\,(s^{\alpha} + |t - t_0|^{\alpha}). \qquad (8)$$

Conversely, $f(.)$ is Lipschitz-α at t_0, $0 \leq \alpha \leq n$, if the two following conditions are verified:

1. There exists $\varepsilon > 0$ and a constant A such that for all points t in a neighborhood of t_0 and any scale s

$$|W_f(s, t)| \leq As^{\varepsilon}. \qquad (9)$$

2. There exists a constant B such that for all points t in a neighborhood of t_0 and any scale s

$$|W_f(s, t)| \leq B\left(s^{\alpha} + \frac{|t - t_0|}{|\log|t - t_0||}\right). \qquad (10)$$

This shows that the WT is particularly efficient to evaluate the local regularity of functions unlike the Fourier transform which is adapted only to global regularity. Indeed, it offers the possibility to analyze the pointwise regularity of a function. This is due to the good localization in the scale-space domain of the wavelet basis. In fact, through the measurement of the decay of $|W_f(s, t)|$ in a two-dimensional neighborhood of t_0 in the scale-space (s, t), one can estimate the Lipschitz exponent, thus stating about the singularity, at the point t_0.

As this method may have, in some cases, a heavy computational load, it is seldom applied in practice. Often, in real-world applications, only the decay of $|W_f(s, t_0)|$ is measured (a one-dimensional neighborhood is handled instead of a two-dimensional one). Nonetheless, it has been shown in [22], through a counter-example, that this practical approach does not always yield reliable results. Thus, a method that allies mathematical tractability and numerical ease was proposed [22].

Theorem 2 ***Lipschitz exponents estimation using wavelet maxima***

Let $\Psi(.)$ be a wavelet with compact support, n vanishing moments and n times continuously differentiable. Let $f(.) \in \mathcal{L}^2(\mathbb{R})$. If there exists a scale $s_0 > 0$ such that for all scales $s < s_0$ and $t \in]a, b[$, $|W_f(s, t)|$ has no local maxima, then for any $\varepsilon > 0$, $f(.)$ is uniformly Lipschitz-n on $]a + \varepsilon, b - \varepsilon[$.

This theorem indicates the presence of a maximum in the modulus of the wavelet transform $|W_f(s, t)|$ at the finer scales where a singularity occurs. Discontinuities in a function f can be assimilated to the fact that $|W_f(s, t)|$ remains constant over a large range of scales in a spatial neighborhood of t_0.

5. Wavelets and Network Attacks

The aforementioned definitions of network attacks are particularly convenient in our context. In fact, all of these attacks can be viewed as abrupt changes in some measurable signals. From mathematical point of view, we represent an attack as a singularity that affects a specific metric. In this section, we use the wavelet-based Lipschitz singularity detection approach presented in the foregoing section in order to detect network attacks. In other terms, anomaly detection can be performed through the detection of singularities in a set of metrics, provided that these metrics are accurately chosen. Obviously, the choice of these metrics is of primal importance as it has a bing influence on the performance of the detection system. However, this issue is out of the scope of this paper.

5.1. *Detecting computer network attacks using the wavelet transform*

In practical situations, the signal of interest $\sigma(.)$ (number of open connections, number of transmitted packets, etc.) is not known for all abscissa t but it is uniformly sampled. Then, what is really handled is a set of values $(\sigma(n))_{n\in\{1,..,N\}}$ where N is the number of samples. As $\sigma(.) \in \mathcal{L}^2(\mathbb{R})$ (it has a finite energy), its discrete wavelet transform can be computed using the sequence $(\sigma(n))_{n\in\{1,..,N\}}$. According to the results discussed in the previous section, the Lipschitz regularity of $\sigma(t)$ can be determined by studying the decay of $\left|W_\sigma(2^{-j}, k2^{-j})\right|$ across the scales j. The following proposition is a straightforward implication of theorem 2.

Proposition 3 ***Characterizing computer attacks using the wavelet transform***

Let $\Psi(.)$ be a wavelet with compact support, n vanishing moments and n times continuously differentiable. Let $\sigma(.) \in \mathcal{L}^2(\mathbb{R})$ be a function representing a monitored metric. If there exist $J \geq 2$ and k_0 such that:

(i) There exists $j_0 \in \{1, .., J\}$, such that $(2^{-j_0}, k_0 2^{-j_0})$ corresponds to a local maxima,

(ii) for all $j \in \{j_0, .., J-1\}$, $\left|W_\sigma(2^{-j-1}, k_0 2^{-j-1})\right|$ is a local maxima,

(iii) for all $j \in \{j_0, .., J-1\}$, $\left|W_\sigma(2^{-j-1}, k_0 2^{-j-1})\right| \geq \left|W_\sigma(2^{-j}, k_0 2^{-j})\right|$,

then, the point k_0 corresponds to a computer attack.

More simply, if a local maxima is detected at a point $(2^{-j_0}, k_0 2^{-j_0})$ and if k_0 corresponds to local maxima with increasing modulus at scales finer than 2^{-j_0}, then k_0 corresponds to a computer attack.

The main advantages of this method are the good spatial localization and the low numerical cost. The first feature is due to the wavelet spatial properties while the second one comes from the fact the wavelet transform of a signal of size N can be computed in $O(N)$ steps according to Mallat's pyramidal algorithm [19].

This proposition can be used to design a classifier system allowing to discard false peaks at the intrusion detection level. The objective is to state whether a peak in the monitored metric actually has been caused by an attack event or not.

The following algorithm, based on Proposition 3, describes the steps that should be implemented by the classifier.

```
algorithm detect_attack
  # State whether peaks in the monitored
  # signal correspond to actual attack
  # events
  begin
    forall $j$ in $\{1, ..J\}$
      if ismaximum($2^{-j}$,$k2^{-j}$)
        begin
          $i = j$;
          while $i$ < $J$-1
            begin
              if ($\neg$ ismaximum($2^{-i-1}$,$k2^{-i-1}$) $\vee$
              ($|W_\sigma(2^{-i-1}, k2^{-i-1})|$ < $|W_\sigma(2^{-i}, k2^{-i})|$)
                then exit;
              $i = i + 1$;
          end
          if $i = J$
          generate_alert($k$);
        end
  end
```

The function `ismaximum(.,.)` tests if $W_\sigma(.,.)$ has a local maxima at the resolution given in the first argument in the point given in the second argument. The function `generate_alert(.)` states that a given value of the metric σ corresponds to a real attack event.

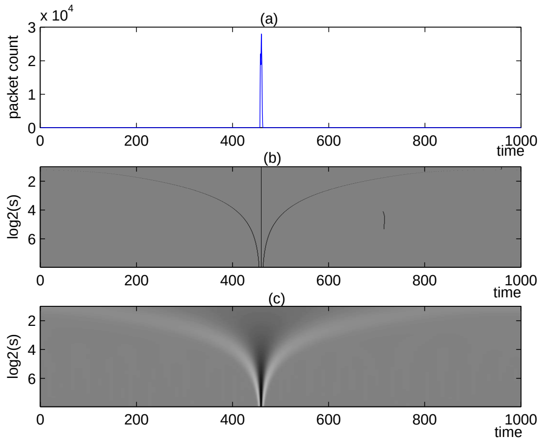

Fig. 2. Analysis of the Lincoln Laboratory DDoS attack scenario (DDoS 1.0).

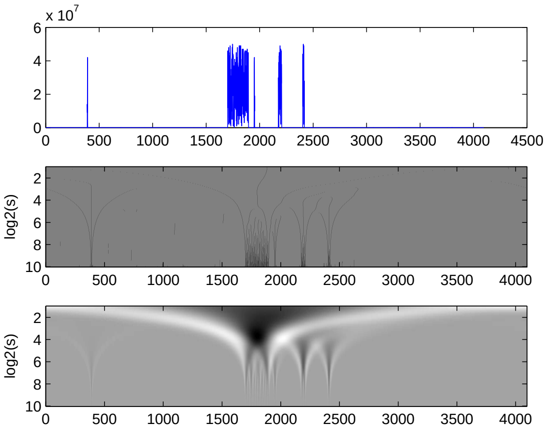

Fig. 3. Packet count per second at the victim side.

5.2. *A Case Study*

In this subsection, we illustrate the result of proposition 3 in a concrete case. We discuss a representative example to demonstrate the benefit of the use of wavelet theory in computer network intrusion detection. We use data corresponding to a real attack provided by the Information System Technology Group of MIT (Massachusetts Institute of Technology)[1]. That is a distributed denial-of-service attack using the *mstream* software and corresponding to the first scenario (Lincoln Laboratory Scenario DDoS 1.0) for year 2000. It was performed over multiple steps including probing, breaking in, installing trojan, and launching a Distributed Denial of Service (DDoS) attack against a US government which IP address was 131.84.1.31.

The measured metric in our case is the packet count $\gamma(t)$ received by the victim machine over observation periods that were fixed to one minute. Figure 2(a) represents the evolution of this function over time (several little fluctuations corresponding to normal traffic can not be seen on the plot because of the great amount of packets received at the moment of the attack). The reader would notice the presence of an important peak that corresponds to the DDoS attack. However, in a real case, the administrator could not determine if this peak corresponds to normal traffic or to an attack. To address this problem, we perform the wavelet decomposition of $\gamma(t)$ using the Daubechies wavelet and we study the evolution of the wavelet coefficient maxima across scales. Figure 2(b) represents the behavior of the maxima positions across scales. The reader can notice that, towards finer scales, the position converges to the instant corresponding to the attack occurrence (samples 458-462). Figure 2 allows to deduce that maxima wavelet coefficients are increasing over scales thus implying that the singularity of interest is an attack.
Figure 3 represents a case where five attacks have been simulated. Attack duration and inter-attack period have been considered as two random processes having respectively uniform and gaussian distributions. Even though maxima values decrease across scales (Figure 3(a)), Figure 3(b) shows that the attack instants still correspond to wavelet maxima at the finest scales, meaning our approach is more robust than measuring the maxima decay.

6. Distributing the monitoring points

6.1. *Extending wavelet theory to* $(\mathcal{L}^2(\mathbb{R}))^p$

The wavelet-based detection approach described in the foregoing sections applies when one or more metrics are used to control a single event related to the same attack. In the case of an infrastructure DDoS attack, multiple key components of the network are simultaneously targeted. Therefore, multiple events can be thought of to detect the occurrence of the attack. For instance, when the victim network consists in an ISP infrastructure, intrusion detection should be performed at the router level. To this purpose, different metrics, corresponding to different events, should be considered, implying the introduction of a suitable wavelet-transform. Supposing that p monitoring points are considered in order to gather information about the transmitted data flows, the global metric that be used by the IDS to state whether an intrusion has occurred or not can be modeled by a signal in $(\mathcal{L}^2(\mathbb{R}))^p$.

In the following we extend the results of the previous sections to the case where $f(.)$ is a function in $(\mathcal{L}^2(\mathbb{R}))^p$.

[1] http://www.ll.mit.edu/IST/ideval/index.html

Definition 4 Multiresolution analysis
A multiresolution analysis of $\mathcal{L}^2(\mathbb{R})^p$ is defined as a set of closed subspaces V_j of $\mathcal{L}^2(\mathbb{R})^p$, $j \in \mathbb{Z}$, verifying the following properties:

(i) $V_j \subset V_{j+1}$
(ii) $v(t) \in V_j \Rightarrow v(2t) \in V_{j+1}$
(iii) $v(t) \in V_0 \Rightarrow v(t+1) \in V_0$
(iv) $\bigcup_{j\in\mathbb{Z}} V_j = \mathcal{L}^2(\mathbb{R})^p$ and $\bigcap_{j\in\mathbb{Z}} V_j = \emptyset$
(v) $\exists\, g(t) \in V_0$ such that $(g(t-k))_{k\in\mathbb{Z}}$ is a Riesz base of V_0 (i.e. $\exists\, A > 0$, $B < +\infty$ s.t. $A \leq \sum_{n\in\mathbb{Z}} |\widehat{g}(\omega + 2\pi n)|^2 \leq B$).

If we assume the approximation of a function f at a resolution j is its projection on V_j, then these properties would have the following significations:

- The first condition means that the information contained in the approximation of a function at a resolution j is necessarily included in the approximation at the resolution $j+1$.
- The second condition is the one that implies scale and dilation invariance. From a function $v(t) \in V_j$ (contains no details or fluctuations at scales smaller than 2^{-j}), the function $v(2t)$ can be obtained by squeezing $v(t)$ by a factor of 2. Thus, this function does not contain details at scales smaller than 2^{-j-1}.
- The third condition corresponds to shift invariance of the spaces V_j. If the approximation of a function at the j^{th} resolution $v_j(t)$ belongs to V_j, so do its translates by integers $(v_j(t-k))_{k\in\mathbb{Z}}$.

The extension of wavelet theory to multivariate signals has been widely discussed in the literature. In their seminal work, Geronimo *et al.* [24] propose a fractal-based wavelet decomposition of multivariate signals. In our work, we considered a simpler componentwise decomposition. Even though it does not rigorously define a multiresolution decomposition of $\mathcal{L}^2(\mathbb{R})^p$, our transform allows to efficiently detect distributed network attacks, as it will be shown in the the following.

According to our transform, a function $f(.) \in \mathcal{L}^2(\mathbb{R})^p$ can be represented at the r^{th} resolution by the expression

$$f(t) = \sum_{k\in\mathbb{Z}} \lambda_{-r,k}\varphi_{-r,k}(t) + \sum_{j=-r}^{j=-1}\sum_{k\in\mathbb{Z}} \gamma_{j,k}\psi_{j,k}(t), \tag{11}$$

where $\varphi_{j,k}(t) = \sqrt{2^j}\Phi(2^j t - k)$ and $\psi_{j,k}(t) = \sqrt{2^j}\Psi(2^j t - k)$, $\forall j \in \{-r, .., -1\}$.

$(\lambda_{-r,k})_{k\in\mathbb{Z}}$ can be seen as the coefficients of the projection of f on V_{-r} and $(\gamma_{j,k})_{j\in\{-r,..,-1\},k\in\mathbb{Z}}$ as those of the projection of f on O_j that verifies the following equation:

$$V_{j+1} = V_j \oplus O_j,\ \forall j \in \{-r, .., -1\}. \tag{12}$$

This means that $(\varphi_{j,k}(t))_{k\in\mathbb{Z}}$ (resp. $(\psi_{j,k}(t))_{k\in\mathbb{Z}}$) is a basis of V_j (resp. O_j) for every $j \in \{-r, .., -1\}$.

As an extension to this reasoning, the function $f_{-r}(t) = \sum_{k\in\mathbb{Z}} \lambda_{-r,k}\varphi_{-r,k}(t)$ can be written as the sum of its projections on V_{-r-1} and O_{-r-1}:

$$f_{-r}(t) = \sum_{k\in\mathbb{Z}} \lambda_{-r-1,k}\varphi_{-r-1,k}(t) + \sum_{k\in\mathbb{Z}} \gamma_{-r-1,k}\psi_{-r-1,k}(t). \tag{13}$$

The following theorem demonstrates that for a function $f(.)$ belonging to $(\mathcal{L}^2(\mathbb{R}))^p$, the wavelet coefficients obtained through componentwise decomposition decrease exponentially with regard to scales (for an appropriately chosen wavelet).

Theorem 5 *Let v be the number of vanishing moments for a wavelet $\psi_{j,k}$, and f . Then the wavelet coefficients given in decay as follows: $\|d_{j,k}\| \leq C_v 2^{-j(v+\frac{1}{2})} \max_{\xi\in I_{j,k}} \|f^{(p)}(\xi)\|$*

Proof For $x \in I_{j,k}$ we write the Taylor expansion of f around $x = k.2^{-j}$.

$$f(x) = \left(\sum_{p=0}^{P-1} f^{(p)}(k.2^{-j})\frac{(x-k.2^{-j})^p}{p!}\right) + f^{(P)}(\xi)\frac{(x-k.2^{-j})^P}{P!},$$

where $\xi \in \left[k.2^{-j}, x\right]$.

Restricting the integral to supp$(\psi_{j,k})$ yields

$$\begin{aligned} d_{j,k} &= \int_{I_{j,k}} f(x)\psi_{j,k}(x)dx \\ &= \left(\sum_{p=0}^{P-1} f^{(p)}(k.2^{-j})\frac{1}{p!}\int_{I_{j,k}} (x-k.2^{-j})^p\psi_{j,k}(x)dx\right) \\ &+ \frac{1}{P!}\int_{I_{j,k}} f^{(P)}(\xi)\frac{(x-k.2^{-j})^P}{P!}\psi_{j,k}(x)dx \end{aligned}$$

We consider the integrals where $p = 0, 1, .., P-1$

$$\begin{aligned} &\int_{k.2^{-j}}^{(k+D-1).2^{-j}} (x-k.2^{-j})^p 2^{\frac{j}{2}}\psi(2^j x - k)dx \\ &= 2^{\frac{j}{2}}\int_0^{D-1}\left(\frac{y}{2^j}\right)^p \psi(y)2^{-j}dy \\ &= 2^{-j(p+\frac{1}{2})}\int_0^{D-1} y^p\psi(y)dy \\ &= 0 \end{aligned}$$

because of the P vanishing moments. Therefore, the modulus of the wavelet coefficient verifies

$$\begin{aligned} \|d_{j,k}\| &= \frac{1}{P!}\left\|\int_{I_{j,k}} f^{(P)}(\xi)\frac{(x-k.2^{-j})^P}{P!}2^{\frac{j}{2}}\psi_{j,k}(x)dx\right\| \\ &\leq \frac{1}{P!}M\int_{I_{j,k}}\left\|\frac{(x-k.2^{-j})^P}{P!}2^{\frac{j}{2}}\psi_{j,k}(x)\right\|dx \\ &= 2^{-j(P+\frac{1}{2})}\frac{1}{P!}M\int_0^{D-1}\|y^p\psi(y)\|\,dy. \end{aligned}$$

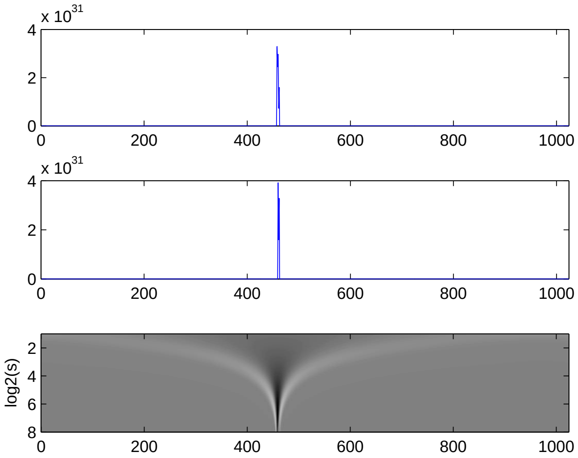

Fig. 4. Wavelet decomposition corresponding to a Distributed Denial of Service attack.

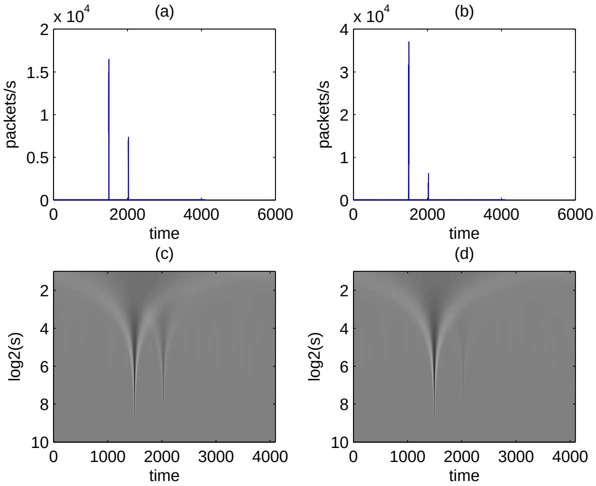

Fig. 5. Wavelet decomposition corresponding to a simulated anomaly.

Defining $C_p = \int_0^{D-1} \|y^p \psi(y)\| \, dy$, where $M = \max_{\xi \in I_{j,k}} \left\| f^{(P)}(\xi) \right\|$we obtain the desired inequality. ■

This theorem shows that even when multiple signals are considered to perform the wavelet transform, the behavior of wavelet coefficients across scales according to Lipschitz regularity do not differ from the case where only one signal is decomposed.

6.2. *Case Study*

We consider again the DDoS attack performed against the experimental MIT DARPA network. We use both the outbound and the inbound DMZ (Demilitarized Zone) packet traffic in order to detect the attack. Those metrics are particularly efficient in the case of a SYN flooding attack because the number of received attack packets (SYN packets) should equal the number of outbound packets (SYN/ACK packets). Figure 4 shows that the decay of the wavelet coefficients still indicates that an attack did occur. It is worth mentioning that in this case, a problem which is frequently encountered is the time synchronization between the multiple signals. Effectively, when gathering metrics at different locations of a specific network, the effect corresponding to the occurrence of an event does not rise at the same instant. Hence, we used the correlation coefficient in order to synchronize the two signals.

Furthermore, we have simulated two distributed flooding attacks attacks having different intensities. Figure 5 shows that the average number of packets per second for the first attack is 1000 while it equals 500 for the second attack(Fig. 5 (a) and (b)). The aforementioned time synchronization mechanism allows to determine the time-shift between the two signals (2). The decay of wavelet coefficients allows to differentiate low-intensity attacks from high-intensity attacks(Figures 5 (c) and (d)). The reader can notice that the first attack is detected around the fifth scale of the wavelet transform while the second attack need a deeper decomposition for being detected.

This example extends the case study of Section 5.2 and demonstrates that our approach allows effectively to differentiate between real and 'fake' attacks. In fact, the simulated attacks, that correspond to peaks in the monitored signals, have not been detected by the wavelet-based approach while they should have generated false alarms if traditional approaches were used. Therefore, according to these case studies that were conducted using real attack data, tracking the evolution of Lipschitz coefficients across scales effectively allows to reduce the false alert rate.

7. Conclusion

In this paper, a method for detecting computer network attacks through the use of wavelet theory and Lipschitz exponents is proposed. The theoretical fundamentals of the method are established and experiments are carried out on traffic containing real attacks to check the correctness of our reasoning. It is shown that the proposed technique allows an efficient detection of computer attacks modulo a good choice of the measured metrics. The merit of our approach is that it exploits the intrinsic properties of the wavelet transform (essentially the low computational cost and the good spatial localization).

Our approach can be extended to detect other types of attacks than DoS and DDoS. Therefore, an interesting issue would be to modify the wavelet-based DDoS attack detection tool in order to differentiate between multiple attacks. A classification scheme can be built to this end.

References

[1] S. Northcuut, J. Novak, ”Network Intrusion Detection,” SAMS, Third Edition, ISBN: 0735712654, 2002.

[2] L. A. Gordon, M. P. Loeb, W. Lucyshyn, R. Richardson, ”10th Annual CSI/FBI Computer Crime and Security Survey,” Computer Security Institute, 2005.

[3] Australian Crime and Security Survey, ISBN: 1-864-99800-8.

[4] Worldwide ISP Security Report, Arbor Networks, September 2005.

[5] G.Carl, G. Kesidis, R. Brooks, S. Rai, ”Denial-of-Service Attack-Detection Techniques,” IEEE Internet Computing, Vol. 10, Issue 1, pp. 82-89, Jan-Feb. 2006.

[6] P. Helman, G. Liepins, “Statistical Foundations of Audit Trail Analysis for the Detection of Computer Misuse,” IEEE Transactions on Software Engineering, Vol. 19, N 9, September 1993.

[7] D. Dasgupta, F.A. Gonzalez, “An intelligent Decision Support System for Intrusion Detection,” Proceedings of the International Workshop on Mathematical Methods, Models and Architectures for Computer Network Security (MMM-ACNS), LCNS, ?ay 21-23, St Petersbourg, Russia.

[8] M.L. Gassata, “A Genetic Algorithm as an Alternative Tool for Security Audit Trail Analysis,” Proceedings of the First International Workshop on Recent Advances in Intrusion Detection, Louvain, Belgium, 1998.

[9] N. Ye , Li X., Chen Q., Emran S., Xu M., Probabilistic Techniques for Intrusion Detection Based on Computer Audit Data, IEEE Transactions on Systems, Man and Cybernetics Part A: Systems and Humans, Vol. 31, No. 4, July 2001.

[10] N. Ye , Emran S., Chen Q., Vilbert S., Multivariate Statistical Analysis of Audit Trails for Host-Based Intrusion Detection, IEEE Transactions on Computers, Vol. 51, No. 7, July 2002.

[11] C. Taylor, J. Alves-Foss, NATE: Network Analysis of Anomalous Traffic Events, a low-cost approach, NSPW01, September 10-13th, 2002, Cloudcroft, New Mexico, U.S.A.

[12] C. Taylor, J. Alves-Foss, An Empirical Analysis of NATE Network Analysis of Anomalous Traffic Events, New Security Paradigms Workshop02, September 23-26, 2002, Virginia Beach, Virginia.

[13] W. DuMouchel, M. Schonlau, A Comparison of Test Statistics for Computer Intrusion Detection Based on Principal Component Regression of Transition Probabilities, Proceedings of 2002 IEEE International Conference on Image Processing (ICIP 2002), vol. 2, pp. 925-928, New York, USA, 2002.

[14] S. Staniford-Chen, L.T. Heberlein, Holding Intruders Accountable on the Internet, Proceedings of the 1995 IEEE Symposium on Security and Privacy, pp. 39-49, Oakland, Canada,1995.

[15] H. Shah, J. Undercoffer, Joshi A., Fuzzy Clustering for Intrusion Detection, Proceedings of the IEEE International Conference on Fuzzy Systems. St. Louis, MO, May 2003.

[16] J.P. Antoine, R. Murenzi, B. Piette, “Image Analysis With 2D Continuopus Wavelet Transform: Detection of Position, Orientation and Visual Contrast of Simple Objects,” Wavelets and Applications, Springer-Verlag, pp. 144-159.

[17] D.L. Donoho, I.M. Johnstone, “Ideal Spatial Adapatation by Wavelet Shrinkage,” Biometrika, N 81, pp. 425-455, 1994.

[18] Y. Meyer, ”Ondelettes et Opérateurs,” Hermann, ISBN: 2 7056 6125 0 - 186,Paris, 1990.

[19] S.G.
Mallat, “A Theory of Multiresolution Signal Decomposition: the Wavelet Representation,” IEEE Transactions PAMI, Vol. 2, n 7, pp. 674-693, 1989.

[20] O. Rioul, “ A discrete-time multiresolution theory unifying octave-band filter banks, pyramid and wavelet transforms”, IEEE Trans. ASSP, June 1990.

[21] S. Jaffard, “Exposants de Holder en des Points Donns et Coefficients d’Ondelettes,” Notes au Compte-Rendu de l’Acadmie des Sciences, France, Vol. 308, Srie I, pp. 79-81, 1989.

[22] S. Mallat, W.L. Hwang, “Singularity Detection and Processing with Wavelets,” IEEE Transactions on Information Theory, Vol. 38, n 2, pp. 617-643, March 1992.

[23] B. Schneier, **“***Secrets and Lies: Digital Security in a Networked World,***”** John Wiley & Sons,ISBN: 0471253111, 2001.

[24] J. Geronimo, D. Hardin, P. R. Massopust, “Fractal Functions and Wavelet Expansions Based on Several Scaling Functions,” J. Approx Theory, vol. 78, pp. 373-401, 1994.

AUTHORS’ BIOGRAPHIES

Mohamed Hamdi Dr. Mohamed Hamdi received his Engineering Diploma, Master Diploma, and PhD in telecommunications from the Engineering School of Communications (Sup’Com, Tunisia) on 2000, 2002, and 2005; respectively. From 2001 to 2005 he has worked for the National Digital Certification Agency (NDCA, Tunisia) where he was head of the Risk Analysis Team. Dr. Hamdi had in charge to build the security strategy for the Tunisian root Certification Authority and to continuously assess the security of the NDCA’s networked infrastructure. He has also served in various national technical committees for securing e-government services. Currently, Dr. Hamdi is serving as a contract lecturer for the Engineering School of Communications at Tunis. He is also member of the Communication Networks and Security Lab (Coordinator of the Formal Aspects of Network Security Research Team), where Dr. Hamdi is conducting research activities in the areas of risk management, algebraic modeling, relational specifications, intrusion detection, network forensics, and wireless sensor networks.

Noureddine Boudriga Prof. Noureddine Boudriga is an internationally known scientist/academic. He received his Ph.D. in Algebraic topology from University Paris XI (France) and his Ph.D. in Computer science from University of Tunis (Tunisia). He is currently a Professor of Telecommunications at University of Carthage, Tunisia and the Director of the Communication Networks and Security Research Laboratory (CNAS) He is the recipient of the Tunisian Presidential award in Science and Research (2004). He has served as the General Director and founder of the Tunisian National Digital Certification Agency. He was involved in very active research and authored or coauthored many chapters and books. He published over 200 refereed journal and conference papers. Prof. Boudriga is the President of the Tunisian Scientific Telecommunications Society.