

Intrusion Detection System based on Matching Pursuit

Rafał Renk, Łukasz Saganowski, Witold Hołubowicz, Michał Choraś
ITTI Ltd., Poznań, Poland;
Adam Mickiewicz University Poznań, Poland; and
Institute of Telecommunications, UT & LS, Bydgoszcz, Poland
rafal.renk@itti.com.pl; luksag@utp.edu.pl; chorasm@utp.edu.pl

Abstract

In this paper we present our original methodology, in which Matching Pursuit is used for networks anomaly and intrusion detection. We propose to use mean projection of the reconstructed network signal to determine if the examined trace is normal or attacked. Experimental results confirm the efficiency of our method.

1 Introduction

Intrusion Detection Systems (*IDS*) are based on mathematical models, algorithms and architectural solutions proposed for correctly detecting inappropriate, incorrect or anomalous activity within a networked systems [1].

Intrusion Detection Systems can be classified as belonging to two main groups depending on the detection technique employed:

1. anomaly detection
2. signature-based detection.

This classification is graphically presented in Figure 1 taken from [1]. Anomaly detection techniques rely on the existence of a reliable characterization of what is normal and what is not, in a particular networking scenario. More precisely, anomaly detection techniques base their evaluations on a model of what is normal, and classify as anomalous all the events that fall outside such a model [2].

If an anomalous behavior is recognized, this does not necessarily imply that an attack activity has occurred: only few anomalies can be actually classified as attempts to compromise the security of the system [3].

In this paper our original methodology for networks anomaly and intrusion detection based on Matching Pursuit is presented. In section 2 the motivation for signal processing methodologies used in intrusion detection is given. In section 3 Matching Pursuit algorithm and base function of

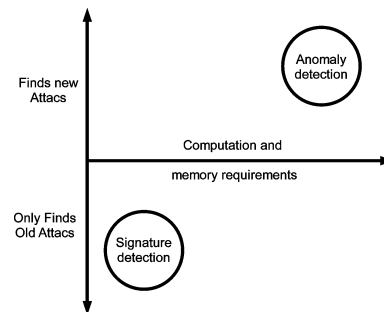


Figure 1. Approaches to Intrusion Detection Systems [source: [1]]

the proposed dictionary design is shown. Experimental results and conclusion are given thereafter.

2 Intrusion Detection Systems based on Signal processing methods

Signal processing techniques have found application in Network Intrusion Detection Systems because of their ability to detect novel intrusions and attacks, which cannot be achieved by signature-based approaches. It has been shown that network traffic presents several relevant statistical properties when analyzed at different levels (e.g. self-similarity, long range dependence, entropy variations, etc.) [4].

Approaches based on signal processing and on statistical analysis can be powerful in decomposing the signals related to network traffic, giving the ability to distinguish between trends, noise, and actual anomalous events. Wavelet-based approaches, maximum entropy estimation, principal component analysis techniques, and spectral analysis, are examples in this regard which have been investigated in the recent years by the research community [5]-[9].

A powerful analysis, synthesis, and detection tool in this field is represented by the wavelets. Indeed, time- and scale-

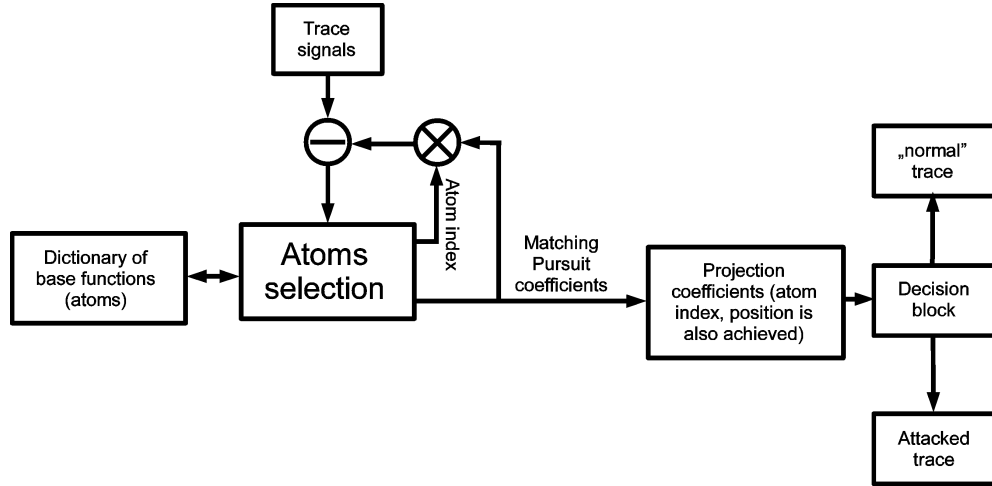


Figure 2. IDS system block diagram

localization abilities of the wavelet transform, make it ideally suited to detect irregular traffic patterns in traffic traces. Recently many wavelet-based methods for detection of attacks have been tested and documented. Some are based on the continuous wavelet transform analysis, most of them however refer to the discrete wavelet transformation and the multiresolution analysis [4].

However, Discrete Wavelet Transform provides a large amount of coefficients which not necessarily reflect required features of the network signals.

Therefore, in this paper we propose another signal processing and decomposition method for anomaly/intrusion detection in networked systems. We developed original Anomaly Detection Type *IDS* algorithm based on Matching Pursuit.

The general overview of our intrusion detection system is presented in Figure 2.

3 Intrusion Detection System based on Matching Pursuit

3.1 Introduction to Matching Pursuit

Matching Pursuit signal decomposition was proposed by Mallat and Zhang [10].

Matching Pursuit is a greedy algorithm that decomposes any signal into a linear expansion of waveforms which are taken from an overcomplete dictionary D . The dictionary D is an overcomplete set of base functions called also atoms.

$$D = \{\alpha_\gamma : \gamma \in \Gamma\} \quad (1)$$

where every atom α_γ from dictionary has norm equal to 1:

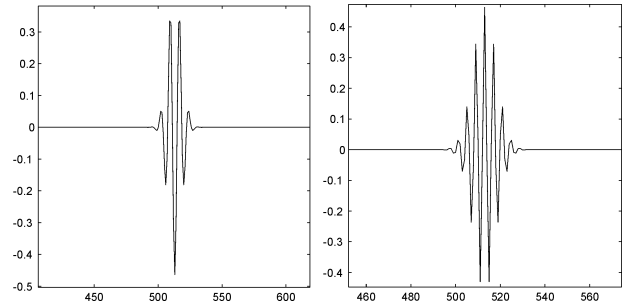


Figure 3. Example dictionary atoms

$$\|\alpha_\gamma\| = 1 \quad (2)$$

Γ represents set of indexes for atom transformation parameters such as translation, rotation and scaling.

Signal s has various representations for dictionary D . Signal can be approximated by set of atoms α_k from dictionary and projection coefficients c_k :

$$s = \sum_{n=0}^{|D|-1} c_k \alpha_k \quad (3)$$

To achieve best sparse decomposition of signal s (min) we have to find vector c_k with minimal norm but sufficient for proper signal reconstruction. Matching Pursuit is a greedy algorithm that iteratively approximates signal to achieve good sparse signal decomposition. Matching Pursuit finds set of atoms α_{γ_k} such that projection of coefficients is maximal. At first step, residual R is equal to the entire signal $R_0 = s$.

$$R_0 = \langle \alpha_{\gamma_0}, R_0 \rangle \alpha_{\gamma_0} + R_1 \quad (4)$$

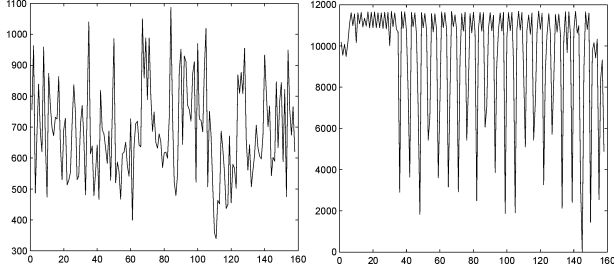


Figure 4. Trace - number of packets per second, normal (on the left) and attacked

If we want to minimize energy of residual R_1 we have to maximize the projection $|\langle \alpha_{\gamma_0}, R_0 \rangle|$. At next step we must apply the same procedure to R_1 .

$$R_1 = \langle \alpha_{\gamma_1}, R_1 \rangle \alpha_{\gamma_1} + R_2 \quad (5)$$

Residual of signal at step n can be written as follows:

$$R^n s = R^{n-1} s - \langle R^{n-1} s | \alpha_{\gamma_k} \rangle \alpha_{\gamma_k} \quad (6)$$

Signal s is decomposed by set of atoms:

$$s = \sum_{n=0}^{N-1} \langle \alpha_{\gamma_k} | R^n s \rangle \alpha_{\gamma_k} + R^n s \quad (7)$$

Algorithm stops when residual $R^n s$ of signal is lower then acceptable limit.

3.2 Base Function Dictionary

In the proposed *IDS* solution 1D real Gabor base function (Equation 8) was used to build dictionary [11]-[13].

$$\alpha_{u,s,\xi,\phi}(t) = c_{u,s,\xi,\phi} \alpha\left(\frac{t-u}{s}\right) \cos(2\pi\xi(t-u) + \phi) \quad (8)$$

where:

$$\alpha(t) = \frac{1}{\sqrt{s}} e^{-\pi t^2} \quad (9)$$

$c_{u,s,\xi,\phi}$ - is a normalizing constant used to achieve atom unit energy,

In order to create overcomplete set of 1D base functions dictionary D was built by varying subsequent atom parameters: Frequency ξ and phase ϕ , Position u , Scale s .

Base functions dictionary D was created with using 10 different scales (dyadic scales) and 50 different frequencies. In Figure 3 example atoms from dictionary D are presented.

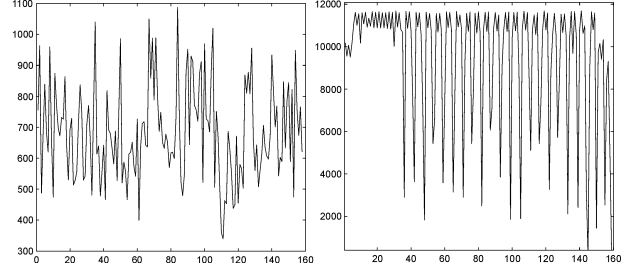


Figure 5. Traces from Figure 4 reconstructed with using Matching Pursuit decomposition. Both signals were reconstructed with 300 base functions

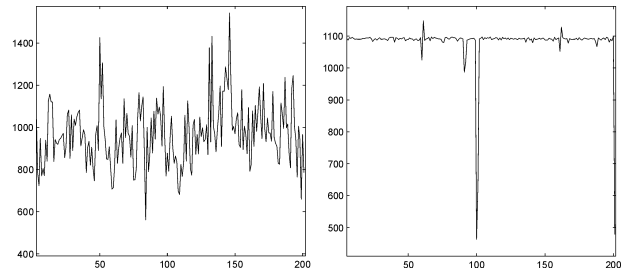


Figure 6. Trace - number of packets per second, normal (on the left) and attacked

4 Experiments and Results

In our experiments we decided to detect anomalous/attacked flows. We tested our algorithms on normal and attacked traces to evaluate if our method is capable of detecting anomalies/intrusions. For each normal trace, we created corresponding attacked trace by injecting simulated UDP flooding attack.

Decision block (Fig. 2) of our system is based on the mean projection values. We calculate difference $Diff$ between attacked and corresponding normal traces according to Eq. 10.

$$Diff = \frac{|NTMP - ATMP|}{\max(NTMP, ATMP)} \quad (10)$$

where $NTMP$ is a mean projection of normal trace and $ATMP$ is mean projection of attacked trace.

If the value of $Diff > 50\%$ our application signalizes an attack.

The sample traces - normal and attacked, and their projection are presented in Figures 4-7.

Sample results of mean projection calculated for normal and attacked traces, respectively, are shown in the Table 1.

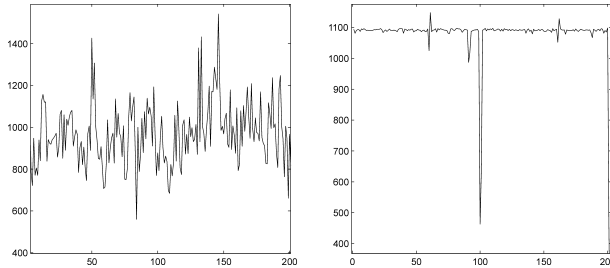


Figure 7. Traces from Figure 6 reconstructed with using Matching Pursuit decomposition. Both signals were reconstructed with 300 base functions

Table 1. Experimental results of intrusion detection based on our matching pursuit methodology.

Trace Number	Normal trace mean projection	Attacked trace mean projection
trace1	42.4	188.6
trace2	38.8	364.6
trace5	9.3	81.7
trace6	9.9	29.6
trace3	60.3	3.7
trace4	45.5	9.9

5 Conclusion

In the article our developments in feature extraction for Intrusion Detection systems are presented. We showed that Matching Pursuit may be considered as very promising methodology which can be used in networks security framework. Upon experiments we may conclude that mean projection differs significantly for normal and attacked traces. Therefore our system easily detects attacked traffic and triggers an alarm.

The major contributions of this paper is a novel algorithm for detecting anomalies based on signal decomposition. In the classification/decision module we proposed to use developed matching pursuit features such as mean projection. We tested and evaluated the presented features and showed that experimental results proved the effectiveness of our method.

6 Acknowledgement

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no.

216585 (INTERSECTION Project).

7 References

1. Esposito M., Mazzariello C., Oliviero F., Romano S.P., Sansone C., Real Time Detection of Novel Attacks by Means of Data Mining Techniques. ICEIS (3) 2005: 120-127.
2. Esposito M., Mazzariello C., Oliviero F., Romano S.P., Sansone C., Evaluating Pattern Recognition Techniques in Intrusion Detection Systems. PRIS 2005: 144-153.
3. FP7 INTERSECTION Project, Deliverable D.2.1: SOLUTIONS FOR SECURING HETEROGENEOUS NETWORKS: A STATE OF THE ART ANALYSIS.
4. FP7 INTERSECTION (INfrastructure for heTERogeneous, Reilient, Secure, Complex, Tightly Inter-Operating Networks) Project Description of Work.
5. C.-M. Cheng, H.T.Kung, K.-S. Tan, Use of spectral analysis in defense against DoS attacks, IEEE GLOBECOM 2002, pp. 2143-2148.
6. P. Barford, J. Kline, D. Plonka, A. Ron, A signal analysis of network traffic anomalies, ACM SIGCOMM Internet Measurement Workshop 2002.
7. P. Huang, A. Feldmann, W. Willinger, A non-intrusive, wavelet-based approach to detecting network performance problems, ACM SIGCOMM Internet Measurement Workshop, Nov. 2001.
8. L. Li, G. Lee, DDos attack detection and wavelets, IEEE ICCCN03, Oct. 2003, pp. 421-427.
9. A. Dainotti, A. Pescape, G. Ventre, Wavelet-based Detection of DoS Attacks, 2006 IEEE GLOBECOM - Nov 2006, San Francisco (CA, USA).
10. S. Mallat and Zhang Matching Pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing.*, vol. 41, no 12, pp. 3397-3415, Dec 1993.
11. J.A. Troop. Greed is Good: Algorithmic Results for Sparse Approximation. *IEEE Transactions on Information Theory.*, vol. 50, no. 10, october 2004 r.
12. R. Gribonval Fast Matching Pursuit with a Multiscale Dictionary of Gaussian Chirps. *IEEE Transactions on Signal Processing.*, vol. 49, no. 5, may 2001.
13. P. Jost, P. Vandergheynst and P. Frossard Tree-Based Pursuit: Algorithm and Properties. *Swiss Federal Institute of Technology Lausanne (EPFL), Signal Processing Institute Technical Report.*, TR-ITS-2005.013, May 17th, 2005.