

A Novel Signal-Based Approach to Anomaly Detection in IDS Systems

Lukasz Saganowski^{1,2}, Michał Choraś^{1,2}, Rafał Renk^{1,3}, and
Witold Hołubowicz^{1,3}

¹ ITTI Ltd., Poznań

rafal.renk@itti.com.pl

michal.choras@itti.com.pl

² Institute of Telecommunications

University of Technology & Life Sciences, Bydgoszcz

luksag@utp.edu.pl

³ Adam Mickiewicz University, Poznań

holubowicz@amu.edu.pl

Abstract. In this paper we present our original methodology, in which Matching Pursuit is used for networks anomaly and intrusion detection. The architecture of anomaly-based IDS based on signal processing is presented. 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 in worm detection scenario. The practical usability of the proposed approach in the intrusion detection tolerance system (*IDTS*) in the INTERSECTION project is presented.

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.

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].

In this paper our original methodology for networks anomaly and intrusion detection based on Matching Pursuit is presented. In Section 2 general overview of the proposed architecture and decision block details are shown. Moreover, the motivation for signal processing methodologies used in intrusion detection

is given. In section 3 Matching Pursuit algorithm and base function of the proposed dictionary design is shown. Experimental results and conclusion are given thereafter.

The major contribution of this paper, is the intrusion/anomaly detection algorithm based on the Matching Pursuit. As to our best knowledge, we have not met any other IDS system based on matching pursuit.

Even though our Matching-Pursuit anomaly detection application is not working in a real time now, it is used in the off-network layer of INTERSECTION Intrusion Detection Tolerance System (*IDTS*).

2 Signal Processing Based Network Anomaly Detection

By profiling the properties of normal network traffic and modeling intrusions or unwanted traffic as anomalies, it is possible to detect the occurrence of such events within reasonable time so to activate reaction and response procedures. Determining the normal behavior model, however, is a difficult task due to the presence of different trends in data, which might be influenced by the time of day, the day of week and seasonal variations.

In our approach we store “normal” traces in a reference database. Normal traces represent traffic from days, we are sure no attacks occurred. These reference traces are compared to current, examined traces. Current traces may be either sniffed from traffic or for experimental purposes may represent old attacks (so that the ground truth is known).

The general overview of our intrusion detection system is presented in Figure 1. The overview of a decision block is explained in Figure 2.

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

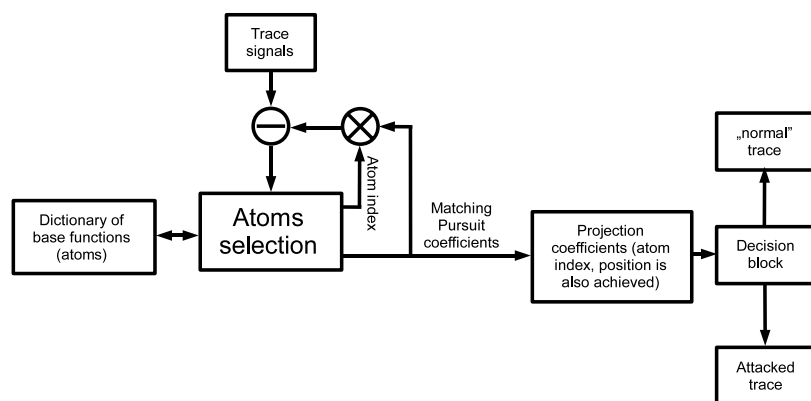


Fig. 1. IDS system block diagram

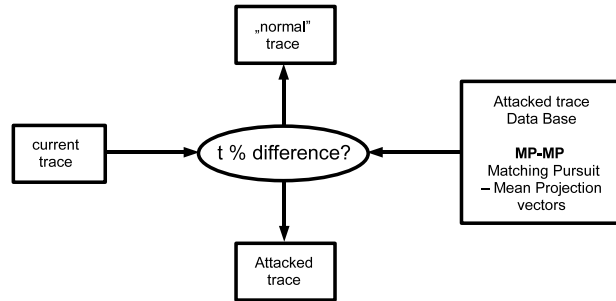


Fig. 2. IDS decision block diagram

different levels (e.g. self-similarity, long range dependence, entropy variations, etc.) [3]. 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 [4]-[8].

A powerful analysis, synthesis, and detection tool in this field is represented by the wavelets. Indeed, time- and scale-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 [3].

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.

3 Anomaly Detection Based on Matching Pursuit

Matching Pursuit signal decomposition was proposed by Mallat and Zhang [9].

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.

$$D = \{\alpha_\gamma : \gamma \in \Gamma\} \tag{1}$$

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

$$\|\alpha_\gamma\| = 1 \tag{2}$$

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 $R_0 = s$.

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

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.

In basic Matching Pursuit algorithm atoms are selected in every step from entire dictionary which has flat structure. In this case algorithm causes significant processor burden. In our coder dictionary with internal structure was used.

Dictionary is built from:

- Atoms,
- Centered atoms,

Centered atoms groups such atoms from D that are as more correlated as possible to each other. To calculate measure of correlation between atoms function $o(a, b)$ can be used [2].

$$o(a, b) = \sqrt{1 - \left(\frac{|\langle a, b \rangle|}{\|a\|_2 \|b\|_2} \right)^2} \tag{8}$$

The quality of centered atom can be estimated according to (9):

$$O_{k,l} = \frac{1}{|LP_{k,l}|} \sum_{i \in LP_{k,l}} o(A_{c(i)}, W_{c(k,l)}) \tag{9}$$

$LP_{k,l}$ is a list of atoms grouped by centered atom. $O_{k,l}$ is mean of local distances from centered atom $W_{c(k,l)}$ to the atoms $A_{c(i)}$ which are strongly correlated with $A_{c(i)}$.

Centroid $W_{c(k,l)}$ represents atoms $A_{c(i)}$ which belongs to the set $i \in LP_{k,l}$. List of atoms $LP_{k,l}$ should be selected according to the Equation 10:

$$\max_{i \in LP_{k,l}} o(A_{c(i)}, W_{c(k,l)}) \leq \min_{t \in D \setminus LP_{k,l}} o(A_{c(t)}, W_{c(k,l)}) \tag{10}$$

In the proposed *IDS* solution 1D real Gabor base function (Equation 11) was used to build dictionary [10]-[12].

$$\alpha_{u,s,\xi,\phi}(t) = c_{u,s,\xi,\phi} \alpha\left(\frac{t-u}{s}\right) \cos(2\pi\xi(t-u) + \phi) \tag{11}$$

where:

$$\alpha(t) = \frac{1}{\sqrt{s}} e^{-\pi t^2} \tag{12}$$

$c_{u,s,\xi,\phi}$ - 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.

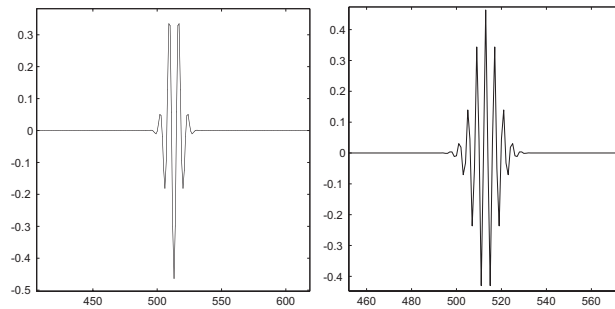


Fig. 3. Example dictionary atoms

4 Experiments and Results

In our experiments we decided to detect worm attacks. We tested our algorithms on normal and attacked traces to evaluate if our method is capable of detecting known worms.

Similarly to the work by Dainotti et al. [13] we tested the efficiency of our algorithms on Slammer and Witty worms. Slammer worm spread in 2003, while Witty spread in March 2004.

In our experiments we use TCP and UDP packets of Slammer and Witty made available by the WIDE-MAWI and CAIDA projects [15][16].

In this paper we will show our algorithm tested on attacked and normal traces.

The attacked traces represent traffic (TCP and UDP packets) from March 20th (Witty) (Figure 4) and March 25th (Slammer) (Figure 5).

The normal traces represent traffic from March 6th and March 13th (Figures 6 7).

The calculated values of Matching Pursuit Mean Projection for our test traces (normal and attacked) are presented in Tables 1-2.

In tables 1 Matching Pursuit Mean Projection values for TCP packets are presented. In tables 2 Matching Pursuit Mean Projection values for UDP packets are given, respectively.

Table 1. Mean Projection values calculated for test TCP traces

TCP Trace	MP
25.03.2004 (Slammer)	620
20.03.2004 (Witty)	667
6.03.2004	453
13.03.2004	373

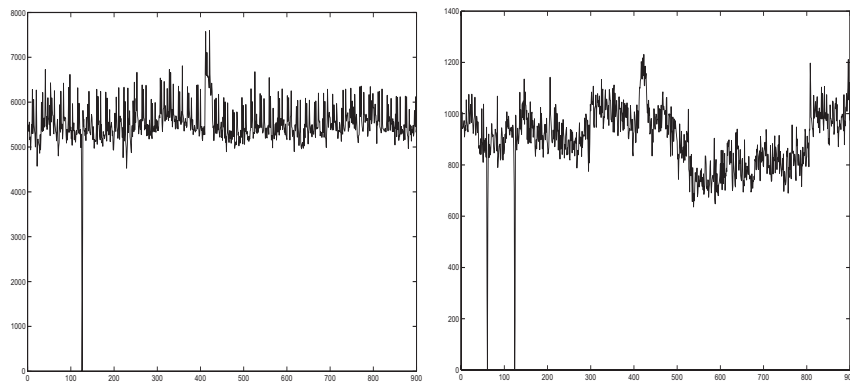


Fig. 4. Traces attacked by Witty worm from March 20th - TCP (left) and UDP (right)

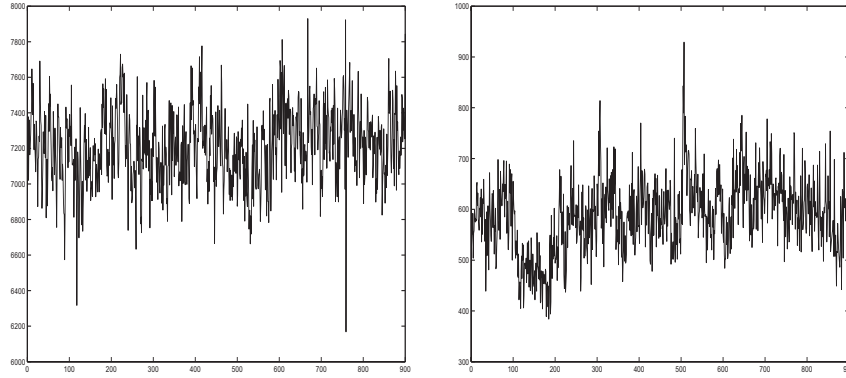


Fig. 5. Traces attacked by Slammer worm from March 25th - TCP (left) and UDP (right)

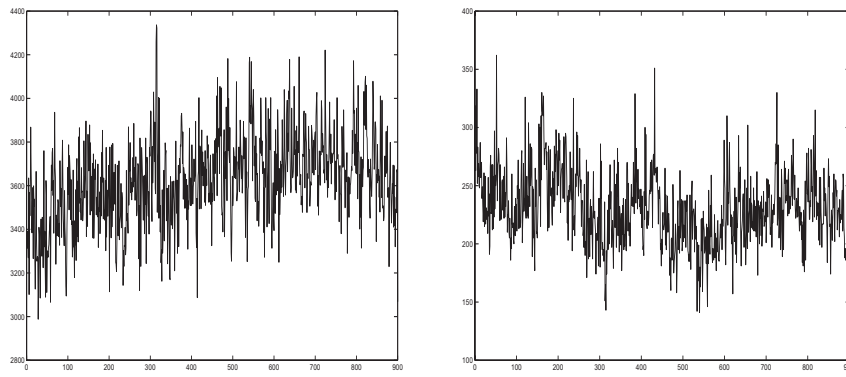


Fig. 6. Normal trace from March 6th - TCP (left) and UDP (right)

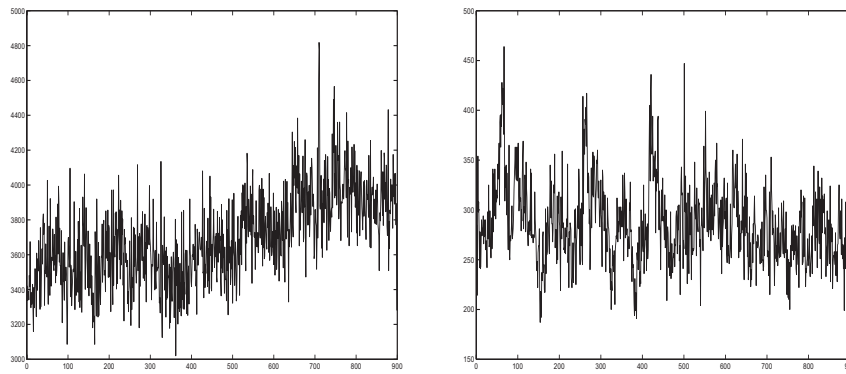


Fig. 7. Normal trace from March 13th - TCP (left) and UDP (right)

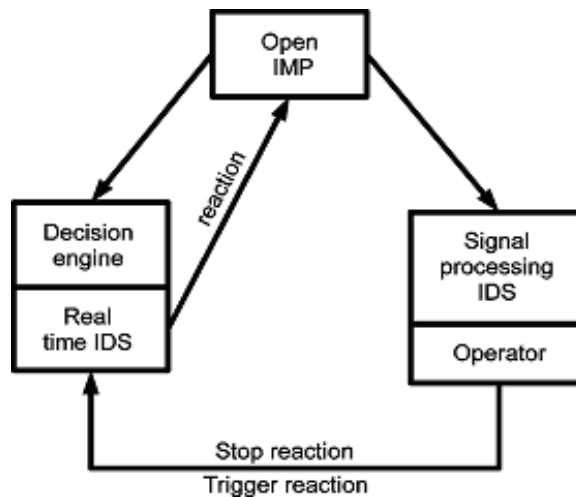
Table 2. Mean Projection values calculated for test UDP traces

UDP Trace	MP
25.03.2004 (Slammer)	82
20.03.2004 (Witty)	127
6.03.2004	32
13.03.2004	40

Decision block of our system is based on the Matching Pursuit Mean Projection values. As presented in Figure 1 we calculate difference *Diff* between examined and normal traces stored in a reference database. If the value *Diff* is larger than a certain threshold t our application signalizes the attack/anomaly.

In the experiments shown here, in the case of Worm attacks, our application was set to $t = 30\%$, which means that if Matching Pursuit Mean Projection differs more than 30% from the reference normal traces the attack should be detected.

As presented in Tables 1-2 mean projection values differ significantly and our IDS application successfully detects Witty and Slammer worms. In our experiments we can report 100% worm detection for TCP and UDP packets with no false alarms. However, so far we tested our method on a limited number of traces. We decided to use known and benchmark traces and worms first. Now we extensively test our method with a larger number of real-networks anonymized traces as well as with the generated traffic traces.

**Fig. 8.** IDS decision block diagram

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 Matching Pursuit Mean Projection differs significantly for normal and attacked traces. Therefore our system successfully detects Slammer and Witty worms.

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.

The proposed Matching Pursuit signal based algorithm applied for anomaly detection IDS will be used as detection/decision module in the INTERSECTION Project security-resiliency framework for heterogeneous networks.

Signal-based anomaly detection type ADS/IDS will be used as the secondary detection/decision module to support real-time IDS. Such approach is proposed for off-network layer of the INTERSECTION framework.

The operator will have a chance to observe the results of signal-based IDS in a near real-time in order to trigger or stop the reaction of real-time IDS. Such approach will both increase the security (less detected anomalies/attacks) and increase the tolerance (less false positives). The overview of the Matching Pursuit IDS role in the INTERSECTION architecture is given in Figure 8.

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).

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), 120–127 (2005)
2. Esposito, M., Mazzariello, C., Oliviero, F., Romano, S.P., Sansone, C.: Evaluating Pattern Recognition Techniques in Intrusion Detection Systems. In: PRIS 2005, pp. 144–153 (2005)
3. FP7 INTERSECTION (INfrastructure for heTEroogeneous, Resilient, Secure, Complex, Tightly Inter-Operating Networks) Project – Description of Work
4. Cheng, C.-M., Kung, H.T., Tan, K.-S.: Use of spectral analysis in defense against DoS attacks. In: IEEE GLOBECOM 2002, pp. 2143–2148 (2002)
5. Barford, P., Kline, J., Plonka, D., Ron, A.: A signal analysis of network traffic anomalies. In: ACM SIGCOMM InternetMeasurement Workshop 2002 (2002)

6. Huang, P., Feldmann, A., Willinger, W.: A non-intrusive, wavelet-based approach to detecting network performance problems. In: ACM SIGCOMM Internet Measurement Workshop (November 2001)
7. Li, L., Lee, G.: DDos attack detection and wavelets. In: IEEE ICCCN 2003, October 2003, pp. 421–427 (2003)
8. Dainotti, A., Pescape, A., Ventre, G.: Wavelet-based Detection of DoS Attacks. In: 2006 IEEE GLOBECOM, San Francisco, CA, USA (November 2006)
9. Mallat, S., Zhang: Matching Pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing* 41(12), 3397–3415 (1993)
10. Troop, J.A.: Greed is Good: Algorithmic Results for Sparse Approximation. *IEEE Transactions on Information Theory* 50(10) (October 2004)
11. Gribonval, R.: Fast Matching Pursuit with a Multiscale Dictionary of Gaussian Chirps. *IEEE Transactions on Signal Processing* 49(5) (2001)
12. Jost, P., Vandergheynst, P., Frossard, P.: Tree-Based Pursuit: Algorithm and Properties. Swiss Federal Institute of Technology Lausanne (EPFL), Signal Processing Institute Technical Report, TR-ITS-2005.013 (May 17, 2005)
13. Dainotti, A., Pescape, A., Ventre, G.: Worm Traffic Analysis and Characterization. In: Proceedings of ICC, pp. 1435–1442. IEEE CS Press, Los Alamitos (2007)
14. Renk, R., Saganowski, L., Hołubowicz, W., Choraś, M.: Intrusion Detection System Based on Matching Pursuit. In: Proc. Intelligent Networks and Intelligent Systems, ICINIS 2008, pp. 213–216. IEEE CS Press, Los Alamitos (2008)
15. WIDE Project: MAWI Working Group Traffic Archive, tracer.csl.sony.co.jp/mawi/
16. The CAIDA Dataset on the Witty Worm - March 19-24, Colleen Shanon and David Moore (2004), <http://www.caida.org/passive/witty>