

---

# Anomaly Detection System based on Redundant Dictionary of Base Functions

Lukasz Saganowski<sup>2</sup>, Michał Choraś<sup>12</sup>, Rafał Renk<sup>13</sup>, and Witold Hołubowicz<sup>13</sup>

<sup>1</sup> ITTI Ltd., Poznań [rafal.renk@itti.com.pl](mailto:rafal.renk@itti.com.pl)

<sup>2</sup> Institute of Telecommunications, University of Technology and Life Sciences, Bydgoszcz [chorasm@utp.edu.pl](mailto:chorasm@utp.edu.pl)

<sup>3</sup> Adam Mickiewicz University, Poznań [name@email.address](mailto:name@email.address)

**Summary.** In this paper innovative recognition algorithm applied to Anomaly Detection System is presented. We propose to use Matching Pursuit Mean Projection (MP-MP) of the reconstructed network signal to recognize anomalies in network traffic. The proposed solutions are used in the security framework within the INTERSECTION Project.

## 1 Introduction

Anomaly Detection Systems can be classified according to:

- the used algorithm
- analyzed features of each packet singularly or of the whole connection
- the kind of analyzed data - whether they focus on the packet headers or on the payload.

Anomaly detection techniques base their evaluations on a model of what is normal, and classify all the events that fall outside such a model as anomalous.

In this paper new solution of ADS system based on signal processing algorithm is presented. ADS analyzes traffic from internet connection in certain point of a computer network. The proposed ADS system uses redundant signal decomposition method based on Matching Pursuit algorithm. ADS based on Matching Pursuit uses Dictionary of Base Functions - *BFD* to decompose input 1D traffic signal (1D signal may represent packets per second) into set of based functions called also atoms. The proposed *BFD* has a ability to approximate traffic signal. Number and parameters of base functions was limited in order to shorten atom search time process.

The paper is structured as follows: first introduction to Matching Pursuit is given in section 2. Then our approach to signal-based ADS is presented (section 3). Finally, in section 4 the performance of presented ADS system was evaluated by set of 40 real traces.

## 2 Introduction to Matching Pursuit

Matching Pursuit signal decomposition was proposed by Mallat and Zhang [1].

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  $\alpha_\gamma$  from dictionary has norm equal to 1:

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

$\Gamma$  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  $\alpha_k$  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  $\alpha_{\gamma_k}$  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:

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

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

### 3 Our Approach to Anomaly Detection Algorithm based on redundant Dictionary of Base Functions

In the 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 ADS system 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 [2]-[4].

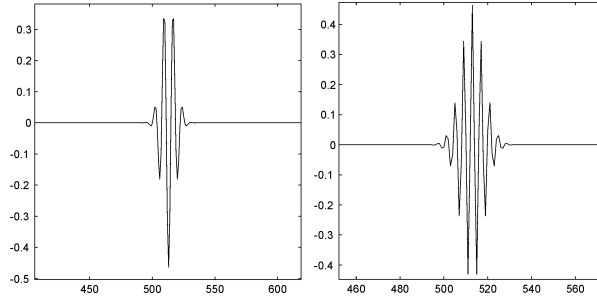


Fig. 1: Example dictionary atoms

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

where:

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

$c_{u,s,\xi,\phi}$  - 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  $\xi$  and phase  $\phi$ , Position  $u$ , Scale  $s$ .

Base functions dictionary  $D$  was created with using 10 different scales (dyadic scales) and 50 different frequencies.

In Figure 1 example atoms from dictionary  $D$  are presented.

In Figure 2 the implementation of Matching Pursuit features (coefficients) in the anomaly detection system is showed.

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 anomaly detection system is presented in Figure 2.

## 4 Experiments and results

Hereby, in our experiments we have evaluated the efficiency of the proposed ADS - Anomaly Detection System. 40 real traces [6][7][8] were used in order to check properties of our ADS algorithm.

The results of our experiments are presented in Figure 3 and Figure 4.

The percentage of the recognized anomalies as a function of encoded atoms from Dictionary of Base Functions is presented in Figure 3. Five dictionaries

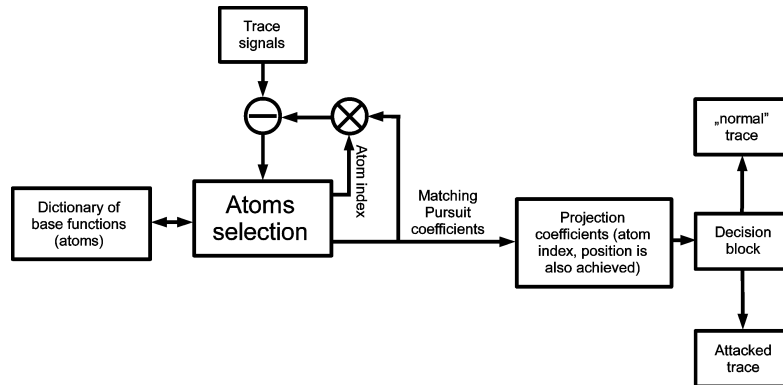


Fig. 2: IDS system block diagram

with different parameters (different number of scales and frequencies) were used. There is no straight relation between size of the dictionary and efficiency of anomaly recognition. Larger dictionary may not improve the effectiveness of anomaly recognition. We have experimentally created the dictionary with parameters allowing for highest percentage of anomaly recognition (Figure 3). After certain threshold (close to 100 atoms) of encoded atoms the percentage of recognized anomalies decreases.

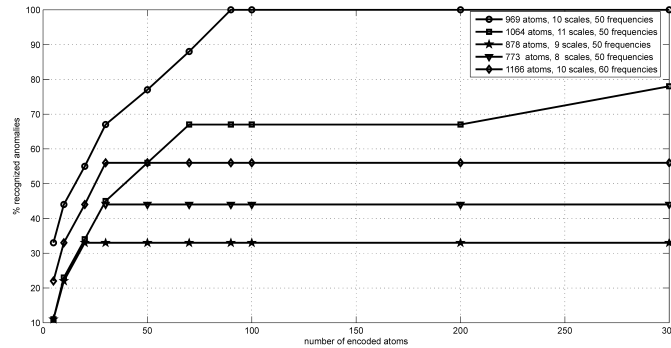


Fig. 3: Percentage of recognized anomalies as a function of encoded atoms

Percentage of the recognized anomalies for Dictionary of Base Functions with approximately constant number of atoms is presented in Figure 4. In this case we try to leave approximately constant number of atoms in dictionary but with different proportions of scales and frequencies.

Trace signals have stochastic nature so it is difficult to create dictionary with sufficient parameters. We created our dictionary experimentally because, so far, there is no universal mathematical recipe how to built dictionary for arbitrary signal .

We have evaluated our ADS system using 40 different real traces. Traces consist of different attacks caused by worms and DDoS (Distributed Denial of Service) programs. Most traces were created from TCP packets (port 80 TCP).

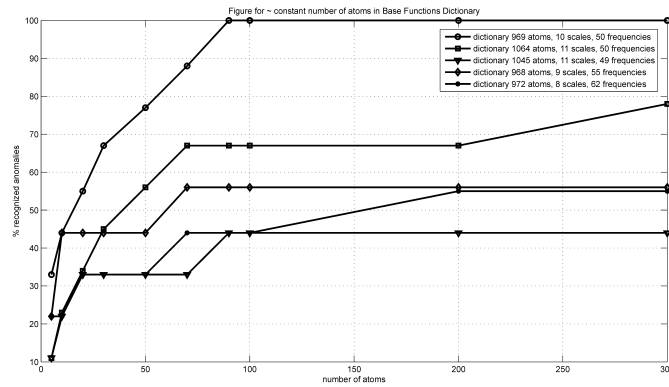


Fig. 4: Percentage of recognized anomalies for Dictionary of Base Functions with approximately constant number of atoms

## 5 Conclusion

In the article our developments in feature extraction for Anomaly Detection Systems are presented. 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 Acknowledgment

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. S. Mallat and Zhang Matching Pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing.*, vol. 41, no 12, pp. 3397-3415, Dec 1993.
2. J.A. Troop. Greed is Good: Algorithmic Results for Sparse Approximation. *IEEE Transactions on Information Theory.*, vol. 50, no. 10, October 2004 r.
3. R. Gribonval Fast Matching Pursuit with a Multiscale Dictionary of Gaussian Chirps. *IEEE Transactions on Signal Processing.*, vol. 49, no. 5, may 2001.
4. 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.
5. Kajan E (2002) Information technology encyclopedia and acronyms. Springer, Berlin Heidelberg New York
6. WIDE Project: MAWI Working Group Traffic Archive at [tracer.csl.sony.co.jp/mawi/](http://tracer.csl.sony.co.jp/mawi/)
7. The CAIDA Dataset on the Witty Worm - March 19-24, 2004, Colleen Shanon and David Moore, [www.caida.org/passive/witty](http://www.caida.org/passive/witty).
8. <http://www.grid.unina.it/Traffic/Traces/ttraces.php>
9. Saganowski L., Choraś M., Renk R., Holubowicz W., A Novel Signal-Based Approach to Anomaly Detection in IDS Systems , M. Kolehmainen et al. (Eds.): ICANNGA 2009, LNCS 5495, pp. 527-536, Springer 2009.
10. L. Coppolino, S. D'Antonio, M. Esposito, and L. Romano, Exploiting diversity and correlation to improve the performance of intrusion detection systems - In Proc of IFIP/IEEE International Conference on Network and Service, 2009.