

Recognizing Anomalies/Intrusions in Heterogeneous Networks

Michał Choraś, Łukasz Saganowski, Rafał Renk, Rafał Kozik,
and Witold Hołubowicz

Abstract. In this paper innovative recognition algorithm applied to Intrusion and/or Anomaly Detection System presented. We propose to use Matching Pursuit Mean Projection (MP-MP) of the reconstructed network signal to recognize anomalies/intrusions in network traffic. The practical usability of the proposed approach in the intrusion detection tolerance system (*IDTS*) in the INTERSECTION project is presented.

1 Introduction

INTERSECTION (INfrastructure for heTErogeneous, Resilient, SEcure, Complex, Tightly Inter-Operating Networks) is a European co-funded project in the area of secure, dependable and trusted infrastructures. The main objective of INTERSECTION is to design and implement an innovative network security framework which comprises different tools and techniques for intrusion detection and tolerance.

The INTERSECTION framework as well as the developed system called *IDTS* (Intrusion Detection and Tolerance System) consists of two layers: in-network layer and off-network layer.

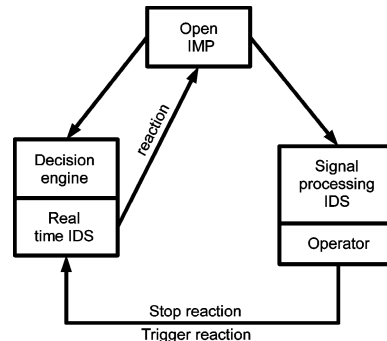
The role of the off-network layer is to support network operators in controlling complex heterogeneous and interconnected networks and real-time

Michał Choraś · Rafał Renk · Witold Hołubowicz
ITTI Ltd., Poznań
michal.choras@itti.com.pl

Michał Choraś · Łukasz Saganowski · Rafał Kozik
Institute of Telecommunications, UT&LS Bydgoszcz
chorasm@utp.edu.pl

Rafał Renk · Witold Hołubowicz
Adam Mickiewicz University, Poznań
holubowicz@amu.edu.pl

Fig. 1 INTERSECTION in-network and off-network approach to intrusion/anomaly detection in complex heterogeneous networks



security processes such as network monitoring, intrusion detection, reaction and remediation.

In this paper we focus on presenting innovative approach to the off-network Anomaly Detection System. Novel techniques applied to ADS/IDS based on signal processing are proposed.

Signal-based anomaly detection type 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 (e.g. at telecoms premises) 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 based IDS/ADS role in the INTERSECTION architecture is given in Figure 1.

2 Intrusion Detection System Based on Matching Pursuit

2.1 Rationale and Motivation

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 [1].

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 [2]-[6]. 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.

2.2 Introduction to Matching Pursuit

Matching Pursuit signal decomposition was proposed by Mallat and Zhang [7]. 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:

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

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 :

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

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.

2.3 Our Approach to Intrusion Detection Algorithm

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].

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

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

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

$LP_{k,l}$ 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:

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

In the proposed *IDS* solution 1D real Gabor base function (Equation 11) was used to build dictionary [8]-[10].

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

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 ξ and phase ϕ , Position u , Scale s [11]. Base functions dictionary D was created with using 10 different scales (dyadic scales) and 50 different frequencies.

3 Experiments and Results

In the following section experimental results are shown. Both real network traces as well as simulated network traces (using NS2) had been used in the verification phase.

3.1 *Experimental Results Based on Real Traces*

In our previous work we presented the usability and performance of our ADS in recognizing normal/attacked traces from Mawi and Caida projects [13]. We also showed efficiency of our method in detecting known worms [14].

Hereby we focused on anomaly detection scenario. Therefore we used traces offered by INTERSECTION projects partners, namely CINI - Unina (University of Napoli) and Fraunhofer, with normal and anomalous traces from real networks [15]. In the tested traces, all anomalies are due to attacks, however normally, it is not always the case (in fact most network anomalies are not dangerous).

We calculated $MP - MP$ for all traces in order to determine if the traffic is normal or anomalous. Firstly, we trained the system with 25% of traces. The remaining traces were used in the testing phase.

In the following tables (Tables 1 and 2) the suspicious traces (with anomalies/attacks) are marked (by bold). The threshold (for allowing MP-MP value difference between "normal" traces and current examined trace) is set to 30%.

3.2 *NS2 Simulation Experiments*

We run network simulations using LAN topology of 10 nodes. Three of them were routers only. Node number 3 is a server which is being attacked by node 4. We measured packed flow on node 2. Of course, simulations were run several times to achieve several vectors.

Table 1 Matching Pursuit Mean Projection for Port 80 TCP test traces

Port 80 TCP	Matching Pursuit Mean Projection
Unina1	211.71
Unina2	89.33
Unina3	255.33
Unina4	170.65
Unina5	186.64
Unina6	285.65
Unina7	285.65
Unina8	212.91
Unina9	339.91
Unina10	393.06
Unina11	277.08
Unina12	476.88
Unina13	309.30
Unina14	242.93
Unina15	234.61

Table 2 Matching Pursuit Mean Projection for Port 80 TCP dump

Port 80 TCP	Matching Pursuit Mean Projection
Tcp trace1	576.06
Tcp trace2	676.30
Tcp trace3	450.72
Tcp trace4	611.40
Tcp trace5	478.50
Tcp trace6	592.15
Tcp trace7	409.31
Tcp trace8	321.93
Tcp trace9	567.63
Tcp trace10	507.98
Tcp trace11	360.73
Tcp trace12	508.00
Tcp trace13	565.65
Tcp trace14	576.23

We simulated DoS attack. In such approach, an attacker floods server with requests, slowing down its performance. We had generated background traffic first, then we generated attacks.

The topology of our simulated network is shown in Fig. 2. $MP - MP$ based features showing anomalies/attacks are presented in Fig. 3. The values of $MP - MP$ for normal and attacked traces are shown in Table 3.

framework. Upon experiments we concluded that Matching Pursuit Mean Projection differs significantly for normal and anomalous/attacked traces.

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.

Acknowledgements. 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.: Evaluating Pattern Recognition Techniques in Intrusion Detection Systems. In: PRIS 2005, pp. 144–153 (2005)
2. 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)
3. Barford, P., Kline, J., Plonka, D., Ron, A.: A signal analysis of network traffic anomalies. In: ACM SIGCOMM Internet Measurement Workshop (2002)
4. 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)
5. Li, L., Lee, G.: DDos attack detection and wavelets. In: IEEE ICCCN 2003, October 2003, pp. 421–427 (2003)
6. Dainotti, A., Pescapé, A., Ventre, G.: Wavelet-based Detection of DoS Attacks. In: 2006 IEEE GLOBECOM, San Francisco, CA, USA (November 2006)
7. Mallat, S., Zhang: Matching Pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing* 41(12), 3397–3415 (1993)
8. Troop, J.A.: Greed is Good: Algorithmic Results for Sparse Approximation. *IEEE Transactions on Information Theory* 50(10) (October 2004)
9. Gribonval, R.: Fast Matching Pursuit with a Multiscale Dictionary of Gaussian Chirps. *IEEE Transactions on Signal Processing* 49(5) (May 2001)
10. Jost, P., Vandergheynst, P., Frossard, P.: Tree-Based Pursuit: Algorithm and Properties. In: Swiss Federal Institute of Technology Lausanne (EPFL), Signal Processing Institute Technical Report. TR-ITS-2005.013 (May 17, 2005)
11. Andrysiak, T., Choraś, M.: Image Retrieval Based on Hierarchical Gabor Filters. *International Journal Applied Mathematics and Computer Science (AMCS)* 15(4), 471–480 (2005)

12. Dainotti, A., Pescapé, A., Ventre, G.: Worm Traffic Analysis and Characterization. In: Proceedings of ICC, pp. 1435–1442. IEEE CS Press, Los Alamitos (2007)
13. 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)
14. Saganowski, L., Choraś, M., Renk, R., Hołubowicz, W.: Signal-based Approach to Anomaly Detection in IDS Systems. *International Journal of Intelligent Engineering and Systems* 1(4), 18–24 (2008)
15. <http://www.grid.unina.it/Traffic/Traces/ttraces.php>

