PENERAPAN METODE *PRINCIPAL COMPONENT ANALYSIS* (PCA) UNTUK DETEKSI ANOMALI PADA JARINGAN *PEER-TO-PEER (P2P) BOTNET*

Adhitya Nugraha[1], Nova Rijati[2]
[1,2]Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang
Jl. Nakula I no. 5 – 11, Semarang, 50131, (024) 3517261
E-mail : adhitya@dsn.dinus.ac.id[1], novaola@yahoo.com[2]

Abstrak

Sejak kemunculan peer-to-peer (P2P) Command and Control (C&C) arsitektur, botnet menjadi lebih kuat dibandingkan sebelumnya. Identifikasi anomali dari P2P botnet sangatlah sulit dilakukan padahal proses tersebut merupakan langkah awal yang sangat penting untuk mengidentifikasi kemungkinan adanya potensi ancaman dari malicious bot dalam jaringan. Hal ini menjadi sulit dikarenakan beberapa perilaku dari fitur botnet sangatlah mirip dengan aktifitas jaringan yang sah.Tujuan dari penelitian ini adalah menemukan anomali yang disebabkan oleh peer to peer (P2P) botnets menggunakan metode PCA. Sebagai tambahan, Euclidean distance digunakan untuk mengkalkulasi anomali indeks sebagai parameter pengukuran dari anomali dalam jaringan. Threshold ditetapkan berdasarkan perhitungan pada training set. Setiap pengujian atas sampel test data akan dibandingkan dengan threshold. Apabila hasil kalkulasi test data berada diatas nilai threshold, maka ini menandakan adanya kemungkinan perilaku abnormal pada jaringan. Hasil menunjukan bahwa model kami mampu memberikan akurasi dan efisiensi komputasi dalam mendeteksi perilaku abnormal dari P2P botnet.

Kata kunci: *botnet, P2P Command and Control, deteksi anomali, PCA*

Abstract

Since the emergence of peer-to-peer (P2P) Command and Control (C&C) architecture, botnet are more resilient to defenses and countermeasures than before. Identifying anomalous of P2P botnet traffic is difficult whereas identifying anomalous traffic is first step that essential for diagnosing event that may indicate the potential threat of malicious bots in the network. It difficult because some traffic features of legitimate P2P traffic are quite similar to the malicious traffic.The aim of this project is to find ways to detect the anomaly that caused by peer to peer (P2P) botnets using PCA. In addition, Euclidean distance are used to calculate anomaly index as measurement parameter of anomaly in network traffic. Threshold are calculated from the training dataset. Every incoming point of test vector will compared to threshold. If the calculation results are beyond the parameters of threshold, then this indicates the possibility of abnormal behavior. The results show that our model is promising in terms of detection accuracy and computational efficiency to detect abnormal behaviour of P2P botnet activity.

Keywords: *botnet, P2P Command and Control, anomaly detection, PCA*

1. PENDAHULUAN

Penggunaan internet saat ini terancam oleh berbagai macam serangan yang dimaksudkan untuk mencuri berbagai data dan informasi berharga demi memperoleh keuntungan pribadi [1][2][3]. Perkembangan akan cara exploitasi kelemahan sistem komputer dan jaringan dengan menggunakan *malware, virus, trojan* dan *botnet* menyebabkan setiap pengguna harus selalu waspada akan serangan tersebut.

Botnet merupakan salah satu malware yang memiliki ancaman paling serius terhadap keamanan internet. Botnet mampu melakukan serangan ilegal

seperti *spam, phishing, click fraud,* pencurian password dan *Distributed Denial of Service (DDoS) attack* [3][4].

Conficker merupakan botnet jenis baru yang telah menerapkan arsitektur *peer-to-peer (P2P) Command and Control (C&C).* Botnet jenis ini tidak memiliki kegagalan sistem yang terpusat, penggunaan port yang dinamis dan *payload* yang terenkripsi sehingga sulit untuk dideteksi. Alasan lain P2P botnet semakin sulit dideteksi adalah beberapa fitur botnet dalam jaringan sangat mirip dengan fitur jaringan yang sah sehingga sangat sulit membedakan mana jaringan sah dan yang tidak [5][6][7][8].

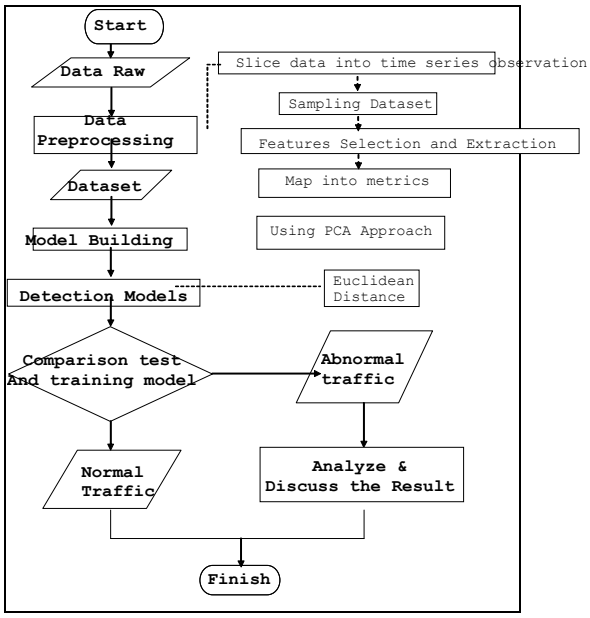
Principal Component Analysis (PCA) merupakan salah satu algoritma machine learning yang biasa digunakan untuk *dimension reduction.* Teknik PCA seringkali digunakan untuk proses *data mining* dan juga *image processing.* Namun, pada tahun 2005, Lakhina et al [9][10][11], memperkenalkan teknik deteksi perilaku pada jaringan untuk menemukan anomali pada jaringan yang disebabkan oleh *Denial of Service Attack* dan *Port Scanning.* Dataset yang digunakan pada penelitian [10] didapatkan dari backbone networks Abilene dan Geant.

Penelitian lain yang dilakukan oleh Wang et al [12][13] mengembangkan model deteksi anomali dengan menerapkan PCA untuk audit *data stream* dalam jumlah besar untuk deteksi secara *real-time.* Penelitian tersebut menunjukan model tersebut memiliki akurasi yang tinggi dan komputasi yang cukup efisien.

Berdasarkan latar belakang tersebut, tujuan dari penelitian ini adalah mengembangkan sebuah model untuk mendeteksi anomali yang disebabkan P2P botnet dengan menggunakan teknik PCA.

2. METODE

Dalam penelitian ini, dibangun sebuah model deteksi anomali P2P botnet dengan menggunakan teknik PCA. Berikut adalah skema yang digunakan dalam penelitian ini.



Gambar 1. *Proposed* Model Deteksi untuk P2P Botnet

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data *testbed* dari lalu lintas jaringan yang berhasil didapatkan dari uji coba lab di University Teknikal Malaysia Melaka (UTeM). Experimen dilakukan di dua (2) hari yang berbeda yang terdiri atas *training data* (*P2P normal traffic*) dan test data (*contain P2P botnet traffic*).

2.2 Pengukuran Performa

Pada penelitian ini, penulis merujuk kepada sistem pengukuran IDS [14] untuk mengukur performa dari sistem yang akan dibuat dimana parameter yang digunakan adalah sebagai berikut.

- True Positif (TP) : benar mengklasifikasi serangan sebagai serangan
- False Positif (FP) : salah

mengklasifikasi lalu lintas normal sebagai serangan
- True Negatif (TN) : benar mengklasifikasi lalu lintas normal sebagai normal
- False Negatif (FN) : salah mengklasifikasi serangan sebagai lalu lintas normal

2.3 Data *Pre-Processing*

Dalam tahap ini, beberapa langkah *pre-processing* dilakukan untuk menyiapkan data agar lebih mudah diproses. Berikut adalah langkah-langkahnya.

a. Menyiapkan data menjadi *time-series information*
b. Sampling data: menyiapkan data training dan test data
c. *Features selection and extraction* : memilih fitur paket flows yang terdiri atas *Src IP add, Dest IP add, Src Port, Dest Port* dan *protocol*
d. Menyiapkan matriks perhitungan:

Tabel 1: Matriks Flows untuk salah satu fitur (IP address)

	t_1	t_2	...	t_n
IP add^1				
IP add^2				
⋮				
IP add^m				

2.4 Pengembangan Model Deteksi Menggunakan PCA

Dalam tahap ini, dikembangkan model deteksi P2P botnet berdasarkan dataset yang sudah dipersiapkan sebelumnya. Hasil dari tahap ini, terbentuk sebuah pola perilaku jaringan pada kondisi normal sebagai model awal untuk deteksi atas anomali yang disebabkan oleh botnet. Kemudian teknik *euclidean distance* digunakan untuk mengukur jarak hasil test data dari model awal yang digunakan. Hasil pengukuran tersebut disimpan sebagai anomali indeks.

3. HASIL DAN PEMBAHASAN

Dalam studi ini, peneliti melakukan pendekatan *passive network monitoring* [15][16]. Teknik ini dilakukan dengan memonitor dan menangkap lalu lintas yang melewati suatu jaringan dan kemudian dilakukan analisis untuk mengidentifikasi keberadaan dan karakteristik botnet.

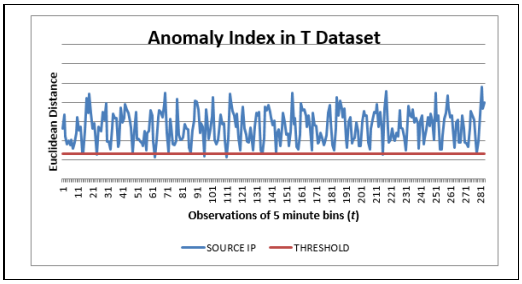
Data yang telah didapatkan terdiri atas normal data dan test data. Tahap pre-processing menghasilkan 284 sampel observasi untuk 5 menit *time-series information.* Fitur yang dipakai untuk analisa adalah 5 *tuples packet flows* seperti yang digunakan pada penelitian Claise [17].

Dari pengolahan data normal didapatkan *threshold* yang akan digunakan sebagai patokan dalam mendeteksi anomali botnet pada test data. Berikut adalah tabel yang menunjukan threshold pada masing-masing fitur analisa.

Tabel 2: Threshold dari normal data

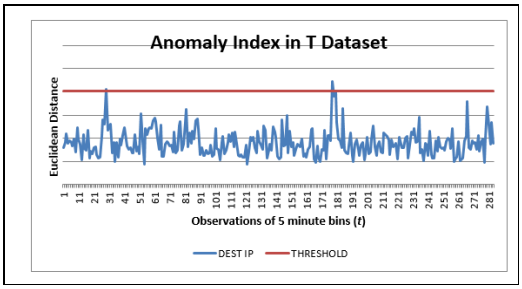
	Five Tuples IP Flow Features				
	IP Source	IP Dest	Port Source	Port Dest	Protocol
Threshold 1	6.64E-22	4.02E-22	1.17E-21	1.53E-21	1.08E-22
Threshold 2	4.66E-21	1.20E-21	3.36E-21	2.33E-21	1.40E-21
Threshold 3	6.92E-21	7.55E-21	6.82E-21	1.94E-21	1.20E-21

Tujuan dari penelitian ini adalah menerapkan PCA model deteksi anomali untuk menemukan pola perilaku yang disebabkan oleh P2P botnet. Dilakukan analisis terhadap test data dimana terdapat anomali yang disebabkan P2P botnet. Dalam hal ini, PCA model digunakan untuk membuktikan adanya anomali pada lalu lintas jaringan yang disebabkan P2P botnet. Berikut adalah hasil penerapan PCA model deteksi untuk 5 fitur *payload* pada jaringan.



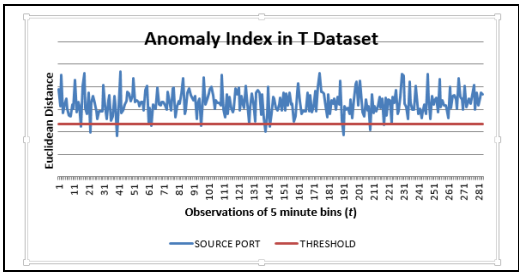
Gambar 2. Anomali indeks pada Source IP

Pada fitur source IP, PCA model mendeteksi 278 sampel sebagai anomali sedangkan 6 sampel sebagai jaringan normal dimana nilai TP adalah 97.8873 % dan FN adalah 0.9615 %.



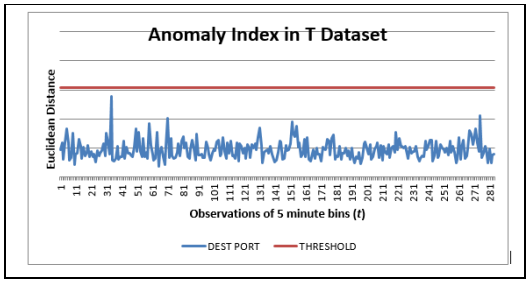
Gambar 3. Anomali indeks pada Dest IP

Pada fitur dest IP, PCA model mendeteksi 2 sampel sebagai anomali sedangkan 282 sampel sebagai jaringan normal dimana nilai TP adalah 0.7042 % dan FN adalah 99.2958 %.



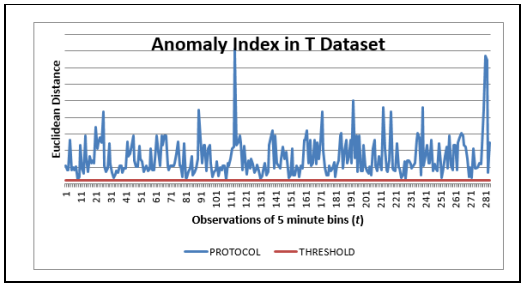
Gambar 4. Anomali indeks pada Source Port

Pada fitur source Port, PCA model mendeteksi 272 sampel sebagai anomali sedangkan 12 sampel sebagai jaringan normal dimana nilai TP adalah 95.7746 % dan FN adalah 4.2254 %.



Gambar 5. Anomali indeks pada Dest Port

Pada fitur dest Port, PCA model mendeteksi 0 sampel sebagai anomali sedangkan 284 sampel sebagai jaringan normal dimana nilai TP adalah 0 % dan FN adalah 100 %.



Gambar 6. Anomali indeks pada Protocol

Pada fitur Protocol, PCA model mendeteksi 0 sampel sebagai anomali sedangkan 284 sampel sebagai jaringan normal dimana nilai TP adalah 100 % dan FN adalah 0 %.

Berikut adalah hasil rangkuman dari seluruh uji coba yang dilakukan terhadap test data.

Tabel 3: Hasil penerapan model PCA pada Test dataset

		TP RATE (%)	FN RATE (%)
Features	IP Source	97.8873	2.1127
	IP Destination	0.7042	99.2958
	Port Source	95.7746	4.2254
	Port Destination	0	100
	Protocol	100	0

4. KESIMPULAN DAN SARAN

Dalam penelitian ini telah diterapkan model deteksi anomali terhadap P2P

botnet dengan menggunakan teknik Principal Component Analysis (PCA). Evaluasi dan hasil yang dilakukan dalam penelitian ini menunjukan hasil yang cukup baik dalam mendeteksi adanya anomali pada jaringan. Beberapa fitur port destination dan IP destination tidak menghasilkan luaran seperti yang diharapkan. Deteksi P2P botnet sangat sulit dilakukan apabila melalui 2 fitur tersebut. Sedangkan protocol menjadi fitur dengan rate keberhasilan terbaik.

Salah satu kendala dalam pendeteksian anomali adalah pada saat pemilihan fitur yang digunakan untuk dianalisa. Kedepannya, kombinasi fitur yang tepat akan meningkatkan hasil yang lebih baik. Dalam penelitian ini, threshold dibangun berdasarkan nilai yang didapat pada normal data. Namun beberapa fitur tampak tidak dapat bekerja dengan baik pada threshold tersebut. Penentuan threshold yang baik kedepannya juga bisa menurunkan nilai false alarm dalam mendeteksi anomali P2P botnet.

DAFTAR PUSTAKA

[1] Feily, M., Shahrestani, A. & Ramadass, S., 2009. *A Survey of Botnet and Botnet Detection*. 2009 Third International Conference on Emerging Security Information, Systems and Technologies, 268-273.

[2] Grizzard, J. et al., 2007. *Peer-to-Peer botnets: Overview and Case Study*. In Proceedings of the first conference on First Workshop on Hot Topics in Understanding Botnets. USENIX Association, p. 1–1.

[3] Karasaridis, A., Rexroad, B. & Hoeflin, D., 2007. *Wide-scale Botnet Detection and Characterization*. 1st Workshop on Hot Topics in Understanding Botnets.

[4] Haq, I. et al., 2011. *What is The Impact of P2P Traffic on Anomaly Detection? In Recent Advances in Intrusion Detection*. Springer, p. 1–17.

[5] Dittrich, D. & Dietrich, S., 2008. *P2P as Botnet Command and Control: A Deeper Insight*. 2008 3rd International Conference on Malicious and Unwanted Software (MALWARE), (June), 41-48.

[6] Piscitello, D., 2010. *Conficker Summary and Review*. ICANN, 1-18.

[7] Ruitenbeek, E.V. & Sanders, W.H., 2008. *Modeling Peer-to-Peer Botnets.* 2008 Fifth International Conference on Quantitative Evaluation of Systems, 307-316.

[8] Schoof, R. & Ralph Koning, 2007. *Detecting peer-to-peer botnets.* University of Amsterdam, 1-7.

[9] Lakhina, A. et al., 2004. *Structural Analysis of Network Traffic Flows.* ACM SIGMETRICS Performance Evaluation Review, 32(1), 61.

[10] Lakhina, A., Crovella, M. & Diot, C., 2004. *Diagnosing Network-Wide Traffic Anomalies*. ACM SIGCOMM Computer Communication Review, 34(4), 219.

[11] Lakhina, A., Crovella, M. & Diot, C., 2005. *Mining Anomalies Using Traffic Feature Distributions*. ACM SIGCOMM Computer Communication Review, 35(4), 217.

[12] Wang, P. et al., 2009. *A Systematic Study on Peer-to-Peer Botnets*. 2009 Proceedings of 18th International Conference on Computer Communications and Networks, 1-8.

[13] Wang, W., Guan, X. & Zhang, X., 2008. *Processing of massive audit data streams for real-time anomaly intrusion detection.* Computer Communications, 31(1), 58-72.

[14] Debar, H., 1999. Towards a *Taxonomy of Intrusion Detection Systems*. Computer Networks, 31(8), 805-822.
[15] Livadas, C. et al., 2006. *Using Machine Learning Techniques to Identify Botnet Traffic*. In In Proceedings of the 2nd IEEE LCN Workshop on Network Security (WoNS2006). Citeseer.
[16] Benferhat, S. & Sedki, K., 2007. *Preprocessing Rough Network Traffic for Intrusion Detection Purposes.* IADIS International Conference, 105-109.
[17] Claise, B., 2008. *Specification of the IP flow Information Export (IPFIX) Protocol for The Exchange of IP Traffic Flow information*. Cisco Systems, Inc, 1-64.