ANOMALY DETECTION PADA INTRUSION DETECTION SYSTEM (IDS) MENGGUNAKAN METODE CLUSTERING ANOMALY DETECTION ON INTRUSION DETECTION SYSTEM (IDS) BY CLUSTERING METHOD

A N A N D A B U D I M U L I A[1], -[2]

[1]Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Intrusion Detection System (IDS) adalah sekumpulan teknik dan metode untuk mendeteksi aktivitas-aktivitas yang terjadi pada level network dan host. Pada sistem ini terdapat dua pendekatan yang dilakukan : signature-based intrusion detection systems dan anomaly detection system. Pendekatan yang pertama memiliki kelemahan yang cukup rentan, yaitu pendeteksian hanya akan dilakukan terhadap data yang sudah didefinisikan. Sementara untuk anomaly detection, selain menggunakan data yang sudah didefinisikan, dapat pula dilakukan dengan menganalisis pola-pola anomali dari paket network yang datang, namun jika salah mengambil parameter maka metode ini justru akan sering mengakibatkan false alarm.

Untuk menganalisis anomaly detection pada paket yang datang dapat dilakukan dengan menggunakan outlier detection scheme. Dengan metode ini, paket-paket yang datang akan dianalisis dengan menggunakan beberapa algoritma, diantaranya adalah clustering. Algoritma clustering pada metode outlier detection scheme melakukan analisis dengan cara meng-cluster-kan data dan menandai cluster terkecil, kemudian cluster terkecil tersebut akan dianggap sebagai anomali.

Dalam Tugas Akhir ini dibangun suatu implementasi pendeteksian intrusion (serangan) terhadap sistem atau jaringan komputer menggunakan metode anomaly detection dengan algoritma cluster-based outlier detection. Proses clustering itu sendiri dilakukan terhadap data koneksi jaringan. Adapun implementasi dilakukan dengan menggunakan bahasa pemrograman HTML, script PHP dan DBMS MySQL.

Pengujian terhadap sistem anomaly detection ini menunjukkan hasil akhir bahwa hasil pendeteksian anomali sangat bergantung pada tiga hal hal, yaitu tergantung pada pemilihan data yang digunakan untuk dianalisis (dataset), jarak maksimal yang diijinkan dari titik pusat cluster atau center ke setiap data yang menjadi anggota dari cluster tersebut atau biasa disebut jari jari cluster, dan perbandingan jumlah data instrusion dengan data normal pada dataset.

Kata Kunci : Intrusion Detection System(IDS), clustering, anomaly detection, outlier detection scheme.

Abstract

Intrusion Detection System (IDS) is a group of techniques and methods for detecting activities that hapenned in network and host level. IDS has two approaches : signature-based intrusion detection system and anomaly detection system. First approach has any weakness, the detection can only done if the intrusion had been definited. Therefore except using the data which had been definited, we can also analyze anomaly patterns from the packets , but if we take the wrong parameter this method could eventually be a false alarm.

Analyze anomaly detection in network data packets can be handled by outlier detection scheme method. With this method we can build the analysis with some algorithms, one of the algorithms is clustering. Clustering algorithm clustered the data and mark the smallest cluster with assumption that smallest cluster as an anomaly.

This final Project will build an implementation of intrusion detection system in computer or network system using anomaly detection method with cluster-based outlier detection algorithm. The process is to clustering data connection record. Implementation use HTML programming language, PHP script, and MySQL DBMS.

Anomaly detection system evaluation shows that the results are depend on three things, data which have been analyzed or data set given and the maximum distance betwen center to each data point that included in that cluster, or cluster radius values and ratio between normal data and instrusion data.

Keywords : Intrusion Detection System(IDS), clustering, anomaly detection, outlier detection scheme.

BAB I
PENDAHULUAN

Berikut ini merupakan sejumlah penjelasan umum mengenai *anomaly detection* pada *Intrusion Detection System* (IDS) yang meliputi latar belakang masalah, rumusan permasalahan, hingga batasan-batasan yang mungkin muncul, serta beberapa penjelasan lainnya.

1.1 Latar Belakang

Seiring dengan semakin berkembangnya teknologi, semakin banyak pula perusahaan yang menjadikan IT sebagai bagian dari proses perancangan pengembangan bisnisnya. Oleh karena itu keamanan sistem menjadi peranan penting untuk menjaga agar sistem yang dibuat tidak rentan terhadap serangan para pengganggu. Meskipun sebuah sistem dibuat sedemikian aman, tapi seiring dengan kemajuan kecepatan *internet* para *cyber attacker* pun semakin meningkat dalam hal jumlah dan kepandaiannya, sehingga sebuah keamanan sistem pun harus terus menerus diperbaharui untuk menanggulangi hal tersebut. Contoh serangan yang cukup berbahaya adalah *code red I* dan *code red II*, *nimda* dan *SQL slammer worm* [2].

Intrusion Detection System (IDS) adalah sekumpulan teknik dan metode untuk mendeteksi aktivitas-aktivitas yang terjadi pada level *network* dan *host* [1,3,4,5]. Terdapat 2 pendekatan pada *IDS* : *signature-based intrusion detection system* dan *anomaly detection system* [1,3]. Pendekatan pertama hanya menganalisis dari *rule-rule* yang telah didefinisikan. Sementara itu pendekatan kedua selain menganalisis dari *rule* yang sudah ada, pendekatan ini juga menganalisis anomali-anomali yang ada pada paket jaringan.

Untuk mendeteksi anomali dapat dilakukan dengan 2 cara, yaitu : *outlier detection scheme dan profiling based technique* [2]. Masing-masing pendekatan ini bisa dilakukan dengan beberapa metode.

Outlier detection scheme adalah teknik mendeteksi anomali dengan cara mengelompokkan data dan mengidentifikasikan data tersebut apakah merupakan deviasi dari sebuah kelakuan normal [5]. Beberapa pendekatan yang bisa dilakukan dengan menggunakan teknik ini adalah : s*tatistic based approach* dan *distance based approach (nearest neighbor approach, clustering based approach* dan *density based approach)* [2,5].

Profiling based technique digunakan pada data *user*, program , dan lainnya yang sudah diprofilkan. Proses *profiling* dapat dilakukan melalui *command line* atau dengan memprofilkan kelakuan *user*.

Diantara kedua teknik diatas, *outlier detection scheme* merupakan teknik yang paling sering digunakan karena metode ini lebih baik dalam hal penanggulangan masalah masalah baru yang terjadi pada jaringan.

Algoritma *clustering* pada metode *outlier detection* menghitung berapa banyak *data point* yang berdekatan dengan *data point* lainnya dengan cara mendefinisikan jari-jari setiap *cluster* dan menentukan jumlah minimal anggota dalam setiap *cluster* agar tidak dianggap sebagai anomali.

1.2 Rumusan Masalah

Dari penjelasan di atas maka dapat dirumuskan beberapa permasalahan pokok, yaitu :

- Bagaimana menentukan apakah paket-paket yang dikirim dapat teridentifikasi sebagai anomali atau tidak dalam sebuah sistem IDS.
- Bagaimana menormalisasi data yang ada sehingga *dataset network connection* dengan atribut bertipe diskrit dan numerik atribut tidak menjadi suatu keterpisahan.
- Bagaimana mengimplementasikan algoritma *clustering* untuk mencari *cluster-cluster* anomali.
- Bagaimana mengevalusi keakuratan hasil prediksi dari sistem ini.

1.3 Tujuan

Berdasarkan permasalahan di atas, tujuan Tugas Akhir ini adalah untuk mengimplementasikan algoritma *clustering* pada metode *outlier detection scheme* dalam mendeteksi anomali yang terjadi pada jaringan.

1.4 Batasan Masalah

Batasan masalah pada Tugas Akhir ini adalah sebagai berikut :

- Pendekatan yang dipakai adalah analisis *anomaly (outlier) detection*.
- Algoritma yang digunakan adalah *cluster-based outlier detection*.

- Data yang akan dianalisis adalah data *network connection record*.
- Data yang dianalisis bersifat data *offline*, yaitu data *training* yang telah dipakai sebagai tesis/penelitian di universitas atau organisasi lain.
- Karena sistem ini mendeteksi anomali, maka metode ini tidak bisa mendeteksi *intrusion* yang dilakukan oleh orang yang memiliki hak akses terhadap *network* (admin) maupun serangan yang dilakukan dengan cara yang terlihat normal.
- *Dataset* yang digunakan tidak boleh didominasi oleh data *intrusion*, ini akan mengakibatkan data *intrusion* yang seharusnya membentuk *cluster* terkecil tersebut menjadi berukuran normal.

1.5 Metodologi

Metodologi yang digunakan dalam memecahkan masalah di atas adalah dengan menggunakan langkah-langkah berikut :

a. Studi literatur

Pada tahap ini dilakukan pengumpulan literature-literatur yang berhubungan dengan topik yang akan dianalisis.

b. Pendalaman materi

Pada tahap ini dilakukan pendalaman materi Tugas Akhir dengan membaca beberapa artikel atau tutorial yang berhubungan dengan *Intrusion Detection System* dan metode *clustering* dalam deteksi anomali.

c. Analisis dan perancangan sistem

Pada tahap ini dilakukan :

- Bagaimana menentukan kelakuan koneksi yang normal maupun tidak.
- Bagaimana cara menentukan atribut yang berguna untuk mendukung proses analisis.
- Bagaimana menormalisasi *dataset* yang ada.
- Kemudian bagaimana menerapkan metode *outlier detection scheme*, khususnya dengan menggunakan algoritma *clustering*.

d. Implementasi dan evaluasi sistem

Pada tahap ini akan dilakukan implementasi dari hasil analisis dan perancangan sistem terhadap metode yang digunakan serta evaluasi apakah sistem ini mampu mendeteksi *intrusion* dengan baik.

1.6 Sistematika Penulisan

Sistematika penulisan dari Tugas Akhir ini adalah sebagai berikut :

BAB I PENDAHULUAN

Menguraikan mengenai latar belakang, rumusan masalah, batasan masalah, tujuan, dan sistematika penulisan yang digunakan.

BAB II DASAR TEORI

Menguraikan berbagai teori mengenai pendekatan yang dapat digunakan pada *Intrusion Detection System*, *outlier detection scheme*, algoritma *clustering*, cara mengevalusi keakuratan hasil *clustering*, dan berbagai teori lain yang mendukung tahap analisis, implementasi dan evaluasi sistem.

BAB III ANALISIS DAN PERANCANGAN

Menguraikan analisis dan perancangan sistem *anomaly detection* pada *Intrusion Detection System*.

BAB IV IMPLEMENTASI DAN EVALUASI

Menguraikan implementasi sistem dan evaluasi terhadap keakuratan hasil deteksi sistem *anomaly detection* dengan algoritma *cluster-based* ini.

BAB V PENUTUP

Berisi kesimpulan dari sistem yang dibuat serta saran untuk pengembangan ke depan dari Tugas Akhir ini.

BAB V
PENUTUP

5.1 Kesimpulan

Berdasarkan hasil analisis dan evaluasi terhadap sistem terdapat beberapa kesimpulan, yaitu :

- Perbandingan jumlah data normal dan data *intrusion* pada proses *clustering* sangat mempengaruhi hasil *clustering* itu sendiri. Semakin kecil jumlah data *intrusion* dibandingkan jumlah data normal maka hasil pendeteksian anomali akan semakin akurat pula.
- Pemilihan *dataset* yang diujikan juga sangat mempengaruhi hasil akhir pendeteksian anomali ini, dimana semakin banyak data *intrusion* yang memiliki kelakukan yang relatif sama dengan kelakukan data normal maka semakin banyak pula *intrusion* yang tidak terdeteksi.
- Penetapan nilai jari-jari *cluster*, *w*, turut mempengaruhi hasil pendeteksian anomali. Dengan penetapan nilai *w* yang kurang tepat maka akan menurunkan nilai *detection rate* dan meningkatkan nilai *false positive*, atau dengan kata lain akan menurunkan keakuratan hasil pendeteksian. Pada Tugas Akhir ini jari jari *cluster* yang paling baik digunakan adalah 0.1

5.2 Saran

Dataset merupakan kunci dari analisis *anomaly detection* pada *instrusion detection system*. Pada Tugas Akhir ini *dataset* yang digunakan terbatas hanya pada satu jenis *dataset*, yaitu data *connection record*. Untuk itu data-data yang berasal dari *instrusion detection system* yang lain tidak dapat langsung digunakan. Oleh karena itu perlu adanya automatisasi untuk semua *dataset* dari berbagai *instrusion detection system* agar analisis yang dikembangkan dapat langsung digunakan untuk berbagai *instrusion detection system*.

DAFTAR PUSTAKA

[1] Aggarwal Charu C, Yu Philip S, *Outlier Detection for High Dimensional Data*

[2] Chen Hulping, *Data Mining approach for intrusion detection system,* 2004

[3] Eskin Eleeazar, Arnold Andrew, Prerau Michael, Portnoy Leonard, Stolfo Sal, *A geometric framework for unsuperviesed anomaly detection : Detection Intrusion in Unlabeled Data*

[4] Fraley C, Raftery A.E, *How Many Clusters? Which cluster method ? Answers Via Model-Based Cluster Analysis*

[5] Han, Kamber, *Data Mining: Concept and Techniques*, DP-3 : 114-116, CA-8 : 335-345, Morgan Kaufmann Publishers, 2001

[6] Lane Terran, Brodley Carla E, *Detecting the Abnormal: Machine Learning in Computer Security, user profiling in IDS*

[7] Lazarević Aleksandar, Srivastava Jaideep, Kumar Vipin, *Data Mining For Intrusion Detection- a tutorial*

[8] Lee Wenke, *A Data Mining Framework for Constructing Features and Models for Intrusion Detection Systems*, Colombia University, 1999

[9] McHugh John, Christie Alan, Allen Julia, *The Role Of Intrusion Detection System*

[10] Rehman Rafeeq Ur, *Intrusion detection system with SNORT*

[11] Tan Pang-Ning, Steinbach Michael, Kumar Vipin, *Intrroduction to Data Mining*, AD-10 : 651-672, Addison Wesley, 2006