ANALISIS METODE *COVARIANCE MATRIX* MENGGUNAKAN TEKNIK LANDMARK WINDOW UNTUK SISTEM DETEKSI ANOMALI TARFIK

ANALYSIS OF COVARIANCE MATRIX METHOD USING LANDMARK WINDOW FOR TRAFFIC ANOMALY DETECTION SYSTEM

Jatmiko Reno Ramadhani[1], Yudha Purwanto[2], Tito Waluyo Purboyo[3]

[1,3]Prodi S1 Teknik Komputer, Fakultas Teknik Elektro, Universitas Telkom
[1]reennno@student.telkomuniversity.ac.id, [2]omyudha@telkomuniversity.ac.id ,
[3]titowaluyo@telkomuniversity@telkomuniversity.ac.id

Abstrak

Dalam beberapa tahun terakhir, fenomena anomali trafik pada lalu lintas jaringan komputer menarik banyak perhatian para peneliti. Menurut [1] serangan *Distributed Denial of Service* (DDoS) adalah jenis serangan yang dapat merugikan trafik jaringan yang sedang digunakan, baik terhadap target serangan maupun seluruh pengguna. Sedangkan peristiwa *flashcrowd* adalah sebuah lonjakan besar pada lalu lintas jaringan internet karena jumlah *user* yang mengakses server naik secara signifikan dan menempatkan tekanan berat pada link jaringan yang mengarah ke server.

Pada tugas akhir ini digunakan teknik statistik *covariance matric* yang tidak mengabaikan fitur satu dengan fitur lainnya, dapat dibuat sistem deteksi anomali dengan mengubah data asli ke ruang fitur *covariance*. Serangan yang ada ini dapat diklasifikasi dengan menggunakan SVM. *Accuracy*, *detection rate* dan *false positive rate* adalah parameter pengujian yang digunakan dalam penelitian.

Hasil dari penelitian ini, algoritma SVM memiliki performansi nilai rata-rata *detection rate* dalam mengklasifikasikan data homogen sebesar 99% pada dataset KDDCUP 99 dan akurasi sebesar 90,5%. Untuk data heterogen performansi menurun dengan meningkatnya nilai FPR pada data yg di uji dengan rata-rata 22,6% karena data diberi noise serangan pada proses preprocessing.

Kata kunci : DdoS, SVM, *flashcrowd*, *covariance matrix*, *Landmark Window*, *detection rate*, *false positive rate*.

Abstract

In recent years, the phenomenon of anomaly traffic on a computer network traffic attracted much attention of research. According to [1] a Distributed Denial of Service (DDoS) is a type of attack that could harm the network traffic that is being used, both againts the target of attacks and all users. While flashcrowd event is a huge spike in network traffic because of the number of internet users who access the server up significantly and put heavy pressure on the network link that leads to the server.

In this final project used statistical techniques covariance matric is not ignore the feature with other features, can be made anomaly detection system by changing the original data into feature space covariance. No attack can be classified by using SVM. Accuracy, detection rate and false positive rate is the testing parameters used in the study.

Results from this study, SVM algorithm has the performance average value in classifying the data detection rate of 99% on a homogenous dataset KDDCUP 99 and an accuracy of 90.5%. For heterogeneous data performance decreased with increasing FPR value at which the data in the test with an average of 22.6% due to the data given in the attack noise preprocessing process.

Keyword : Anomaly detection, covariance matrix, landmark window, SVM, DDoS, flashcrowd, detection rate, false positive rate.

1. Pendahuluan

Pada perkembangan teknologi komputer seperti internet sekarang, keamanan merupakan aspek penting dari suatu sistem. Saat ini hampir seluruh kalangan masyarakat dapat menggunakannya untuk mendapatkan informasi yang luas dan beragam dari seluruh dunia. Banyak kalangan sering kali tidak bertanggung jawab dalam menggunakan teknologi internet saat ini, yang sering kali menyebabkan kerugian. Hal ini pula yang menyebabkan munculnya serangan-serangan di dalam suatu jaringan komputer yang tentunya merugikan. Serangan yang terjadi ini bisa disebut sebagai anomali trafik dimana dapat terjadi *flash-crowd* atau karena serangan *flooding* trafik seperti *Denial of Service* (Dos) dan *Distributed Denial of Service* (DDoS).

Denial of Service (DoS) dan *Disributed Denial of Service* (DDoS) merupakan bentuk serangan *flooding* yang berusaha membuat suatu *host* atau *service* menjadi tak dapat diakses oleh user yang berhak. Sasaran serangan oleh DoS/DDoS adalah *link/bandwidth* untuk membuat sumber daya bandwidth penuh dan sumber daya komputasi pada server agar sistem pengolah kehabisan sumber daya yang berujung oleh jaringan *down* atau *crash* [1]. Sedangkan *flashcrowd* adalah kejadian yang tidak dapat diprediksi tetapi akan terjadi peningkatan akses secara

dramatis/tinggi ke suatu server karena suatu kejadian seperti bencana alam, peluncuran produk, *breaking news*, dll. Metode deteksi menggunakan anomali trafik dikembangkan karena dapat mendeteksi serangan baru yang tidak sama dengan pola normal.

Dalam mendeteksi dan mengatasi serangan di jaringan komputer, dikenal dengan istilah *Intrusion Detection System (IDS).* Pada *Intrusion Detection System (IDS)* dikenal 2 metode yang sering digunakan yaitu *intrusion signature* dan *traffic anomaly based* yang berfungsi untuk mengenali serangan yang terjadi [2]. Pada Tugas Akhir ini penulis menggunakan *traffic anomaly based* untuk mendeteksi serangan dimana keunggulannya tidak memerlukan *database* serangan. Kekurangan dari *traffic anomaly based* ini terletak pada tingkat kesalahan (*false positive rate)* tinggi jika tidak dibuat dengan baik, maka untuk mengatasi kekurangan tersebut kami menggunakan data mining teknik statistik [3].

Dalam pengembangannya sistem deteksi anomali banyak pendekatan untuk mengetahui pola trafik normal sebagai acuan deteksi anomali trafik. Penelitian Tugas Akhir ini menggunakan teknik statistik *covariance matrix* yang tidak mengabaikan hubungan dan ketergantungan antar fitur yang dapat menyebabkan kesalahan deteksi. Deteksi anomali trafik menggunakan *covariance matrix* menggunakan metode *landmark window* bertujuan untuk memperoleh nilai *detection rate* yang tinggi dan *false positive rate* yang rendah.

2. Dasar Teori

2.1 Deteksi Anomali Trafik

Dalam mendeteksi dan mengatasi serangan yang sering terjadi pada jaringan komputer dikenal dua istilah *Intrusion Detection System (IDS)* dan *Intrusion Prevention System (IPS)* [2]. Pada IDS/IPS dikenal dua metode yaitu *intrusion signature* dan *traffic anomaly based*. Untuk *intrusion signature* memiliki kelebihan yaitu *false positive rate* yang rendah selama *database* terus terupdate. Sedangkan untuk *traffic anomaly based* bisa mendeteksi serangan dengan tidak membutuhkan *database* yang harus terus di *update* seperti yang ada pada *intrusion signature*. Pada penelitian [4] teknik deteksi anomali mempunyai kelebihan untuk mengidentifikasi serangan yang tidak diketahui. Tetapi, terdapat masalah besar dalam menentukan pola trafik normal, dimana trafik normal bersifat stokastik, dapat berubah-ubah tergantung dari perilaku pengguna, penempatan dalam jaringan, skala, waktu, dan lain-lain. Anomali trafik dikenal untuk menentukan serangan ketika perilaku subjek yang diamati menunjukkan variasi yang signifikan dari keadaan normal.

2.2 Distribution Denial of Service (DdoS)

Serangan DDoS melibatkan sebagian besar penyerangan yang ditujukan untuk satu target, karena itu DoS untuk user pada satu target sistem. Untuk melancarkan serangan DDoS seorang penyerang menggunakan *botnet* dan sebuah serangan sistem tunggal. Menurut sumber [1] Botnet adalah sebuah aplikasi perangkat lunak yang dapat menjalankan tugas secara otomatis dalam internet dan menampilkan tugas yang berulang ulang seperti index search engine, dan web spidering. Botnet adalah sebuah jaringan besar dari sistem yang berbahaya dan dapat digunakan oleh penyusup untuk membuat serangan DDoS.

2.3 Flashcrowd

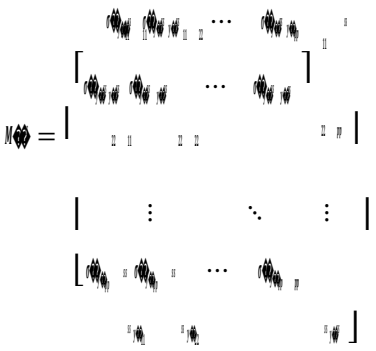
Dalam penjelasan [6] peristiwa *flash crowd* adalah sebuah lonjakan besar pada lalu lintas jaringan internet karena jumlah *user* yang mengakses server naik secara signifikan dan menempatkan tekanan berat pada link jaringan yang mengarah ke server. Kejadian yang sering terjadi yang dapat menyebabkan *flash crowd* adalah bencana alam, *breaking news*, iklan barang, dan sebagainya. Paper [7] telah dipublikasikan analisis beban kerja web dimana lalu lintas atau trafikya sangat padat yaitu pada web piala dunia tahun 1998. Pada dataset piala dunia 1998 menyajikan beban puncak yang menunjukkan banyak klien berulang kali mengunjungi situs dengan waktu antar sesi sangat singkat.

2.4 Covariance Matrix

Covariance matrix digunakan sebagai detector anomali untuk mengubah data ke ruang fitur baru, yang disebut ruang fitur covarian. Penggunaan landmard window untuk menentukan jumlah group data yang membentuk covariance matrix untuk membuktikan efektifitas pada parameter akurasi deteksi. Pada penelitian [9] covariance matrix mengambil sampel secara berurutan untuk mendeteksi beberapa serangan yang terjadi pada traffic network. Pada penelitian [7] untuk membangun sebuah ruang fitur covariance dimana terdapat perbedaan korelasi antara sampel berurutan dan dievaluasi sesuai sampel yang ada.

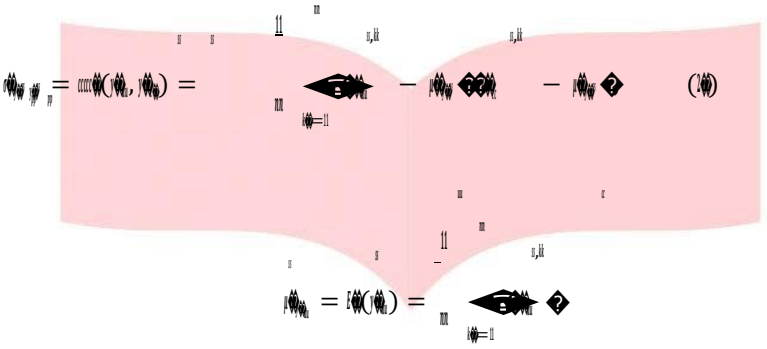
Masalah utama dalam deteksi yaitu selalu mengabaikkan informasi korelasi yang terdapat pada jaringan yang menyebabkan kegagalan dalam efektifitas deteksi. Asumsi *p* fitur fisik $y_1, \ldots, y_p$ untuk menjelaskan bahwa dalam mendeteksi serangan *covariance matrix* tidak mengabaikkan korelasi antar fitur pada pendekatan statistik. Menggunakan *landmark window* untuk observasi W_i, dan akan mendapatkan $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ pada n observasi.

Covariance matrix dari n sampel dilambangkan dengan Σ dan dapat dihitung dengan formula sebagai berikut:



(1)

dimana



(2)

Covariance matrix menjelaskan status jaringan untuk mengukur korelasi antar fitur $y_1, \ldots, y_p$. *Covariance matrix* memiliki kontruksi, dan yang digunakan adalah nilai-nilai kovarian yang terdapat di atas matrix diagonal. Alasannya adalah karena nilai diagonal merupakan nilai varian dari salah satu fitur yang tidak memiliki informasi mengenai korelasi antar fitur.

2.5 Support Vector Machine (SVM)

Support Vector Machine (SVM) pertama kali di perkenalkan oleh Vapnik in 1992 sebagai konsep unggulan dalam bidang pattern recognition, dan dewasa ini merupakan salah satu tema yang berkembang dengan pesat. Sebagai salah satu metode pattern recognition usia SVM masih terbilang muda, namun evaluasi kemampuannya dalam aplikasinya menempatkan sebagai state of the art dalam pattern recognition. SVM adalah metode machine learning yang bekerja atas prinsip Structural Risk Minimization (SRM) dengan tujuan menemukan hyperplane terbaik yang memisahkan dua class pada input space.

Sebuah hyperplane dapat ditulis sebagai :

$$W \cdot X + b = 0 \qquad (3)$$

Dimana W adalah bobot vektor, yaitu, $W = \{w_1, w_{2, \ldots,} w_n\}$, n adalah jumlah atribut, dan b adalah skalar yang sering disebut sebagai bias. Sebagai contoh kita mempertimbangkan dua masukan atribut A_1 dan A_2. Misalnya, $X = (x_1, x_2)$, dimana x_1 dan x_2 adalah nilai atribut A_1 dan A_2, masing-masing untuk X. jika b sebagai bobot tambahan, w_0, kita dapat menulis ulang hyperplane pemisah sebagai :

$$w_0 + w_1 x_1 + w_2 x_2 = 0 \qquad (4)$$

Dengan demikian, setiap titik yang terletak di atas hyperplane pemisah memenuhi :

$$w_0 + w_1 x_1 + w_2 x_2 > 0 \qquad (5)$$

Dengan demikian pula, setiap titik yang terletak di bawah hyperplane pemisah memenuhi :

$$w_0 + w_1 x_1 + w_2 x_2 < 0 \qquad (6)$$

Bobot dapat disesuaikan sehingga hyperplanes dapat mendefinisikan “sisi” dari margin yang ditulis sebagai :

$$H_1 : w_0 + w_1 x_1 + w_2 x_2 \geq 1 \text{ untuk } y = +1 \qquad (7)$$

$H_2: w_0 + w_1 x_1 + w_2 x_2 \geq 1$ untuk $y = -1$ (8)

Artinya, setiap tupel yang jatuh pada atau di atas H_1 milik kelas +1 dan yang jatuh di bawah H_2 milik kelas -1. Menggabungkan persamaan H_1 dan H_2 diperoleh :

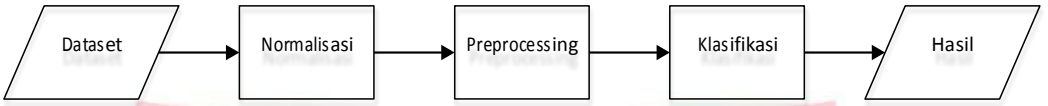
$y(w_0 + w_1 x_1 + w_2 x_2) \geq 1,$ (9) (rumus tidak terbaca)

Setiap sample yang jatuh pada hyperplane H_1 dan H_2 disebut sebagai support vector. Pada dasarnya, vektor dukungan adalah sample yang paling sulit untuk diklasifikasikan tetapi merupakan sumber informasi yang penting tentang klasifikasi [7].

3. Pembahasan

3.1. Deskripsi sistem

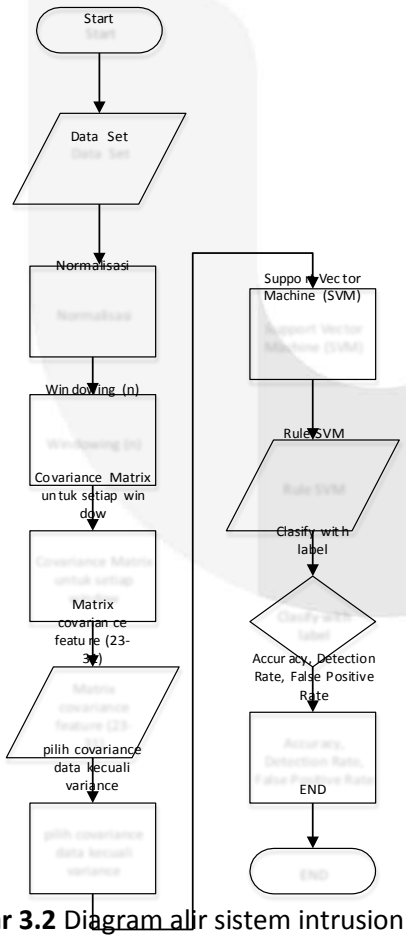
Proses perancangan sistem dalam deteksi anomali trafik yang dilakukan dalam tahap penelitian sebagai berikut. Dataset akan di normalisasi untuk di samakan nilai yang dimiliki dalam dataset. Kemudian pada tahap kedua data di *prepocessing* sehingga dapat diproses menggunakan teknik pembelajaran yang digunakan. Pada proses *preprocessing* dataset yang sudah dinormalisasi akan di hitung menggunakan *covariance matrix* agar tidak ada fitur yang terabaikkan dari dataset yang digunakan.



Gambar 3.1 Pemodelan sistem

3.2. Preprocessing

Data *preprocessing* merupakan strategi agar data cocok digunakan untuk proses pengujian. *Preprocessing* dari pengujian ini yaitu mengolah dataset KDDCUP 99 yang hampir selruhnya berasal dari data DARPA 98, menggunakan *covariance matrix* menggunakan landmark window. KDDCUP 99 merupakan datastandar yang sering digunakan untuk mengevaluasi algoritma untuk pendeteksian intrusi. Seteleh melewati tahap preprocessing kemudian akan dibagi menjadi data training dan data testing. Dengan didapatkannya model SVM dari data training maka bobot nilai yang didapatkan pada pelatihan akan digunakan untuk proses testing sehingga menghasilkan akurasi yang baik. Jika proses *preprocessing* ini tidak dibuat dengan benar akan mengakibatkan *false positive rate* yang tinggi saat di deteksi. Maka dari itu dibuatlah karakteristik fitur-fitur agar dapat menjadi acuan dalam mendeteksi serangan yang terjadi. Berikut diagram alir pada Gambar 3.2.

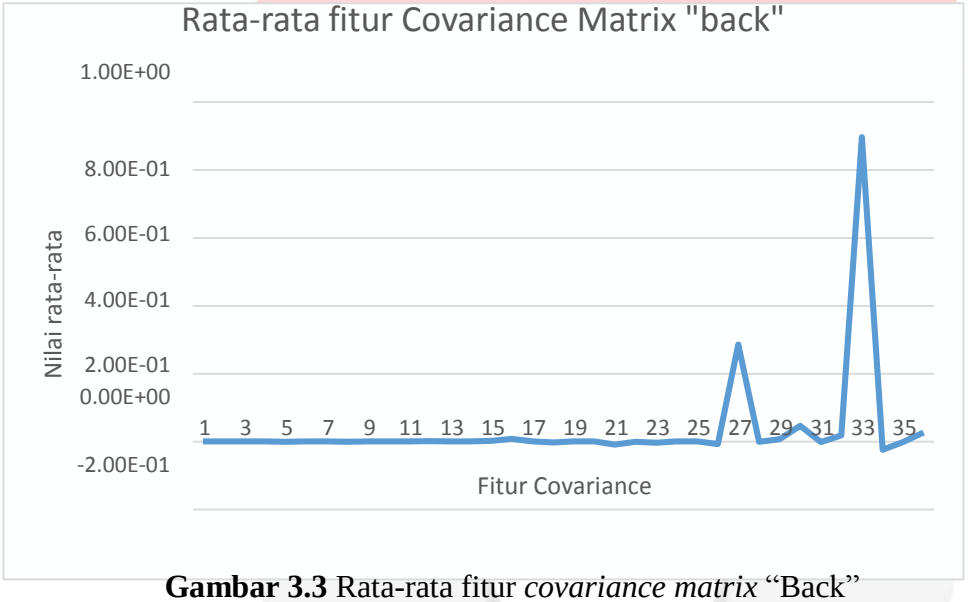


Gambar 3.2 Diagram alir sistem intrusion detection

3.3. Covariance Matrik dari setiap window

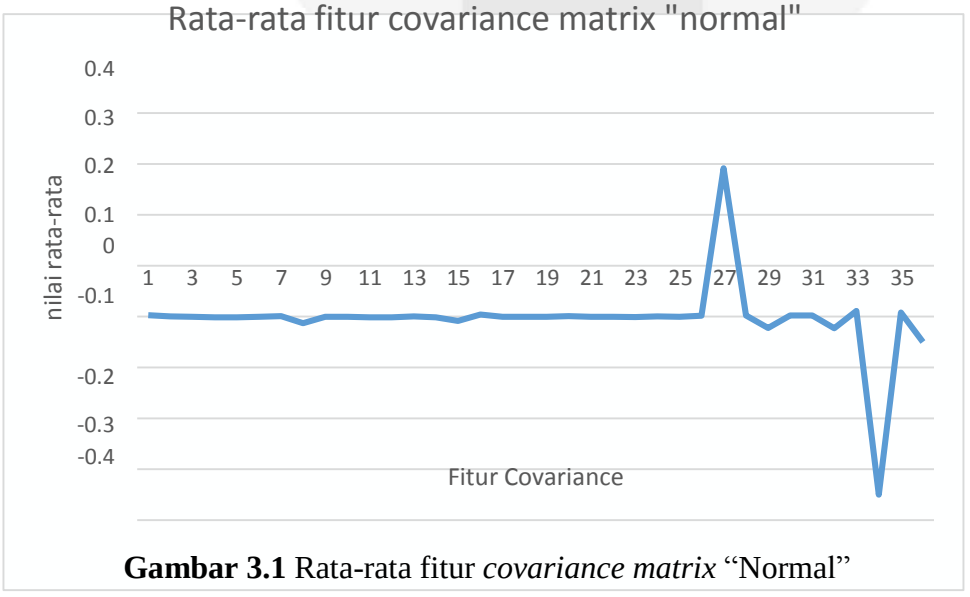
Data Data asli dirubah ke ruang fitur baru yang disebut dengan ruang fitur kovarian, dimana korelasi antar fitur trafik tidak di abaikkan untuk meningkatkan performansi baik untuk diolah lebih lanjut dalam data mining yang dapat menyebabkan efektivitas deteksi menurun [4]. *Covariance matrix* memiliki konstruksi. Nilai *covariance matrix* yang dipakai adalah nilai yang terdapat di atas matrix diagonal. Berikut ditampilkan bentuk trafik seteleh dilakukan *preprocessing* normalisasi dan *covariance matrix*.

Pada Gambar 4.1 ditunjukkan histori 36 fitur trafik *covariance matrix* "Back" dari hasil preprocessing *covariance matrix* untuk setiap window yang di proses. Lalu setiap fitur di hitung rata-ratanya pada Gambar 4.2 agar dapat terlihat lonjakan data.



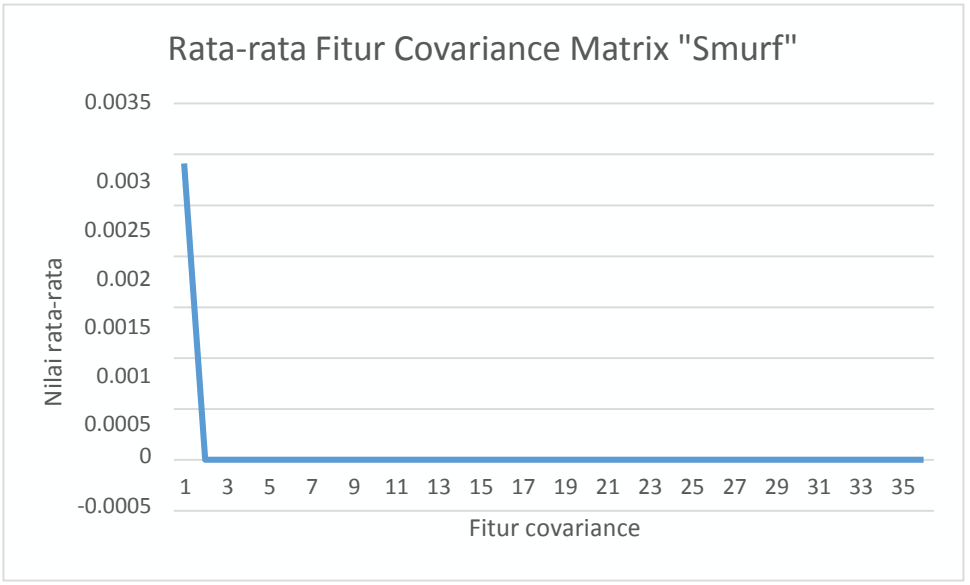
Gambar 3.3 Rata-rata fitur *covariance matrix* "Back"

Pada Gambar 4.3 ditunjukkan histori 36 fitur trafik *covariance matrix* "Normal" dari hasil preprocessing *covariance matrix* untuk setiap window yang di proses. Lalu setiap fitur di hitung rata-ratanya pada Gambar 4.4 agar dapat terlihat lonjakan data.



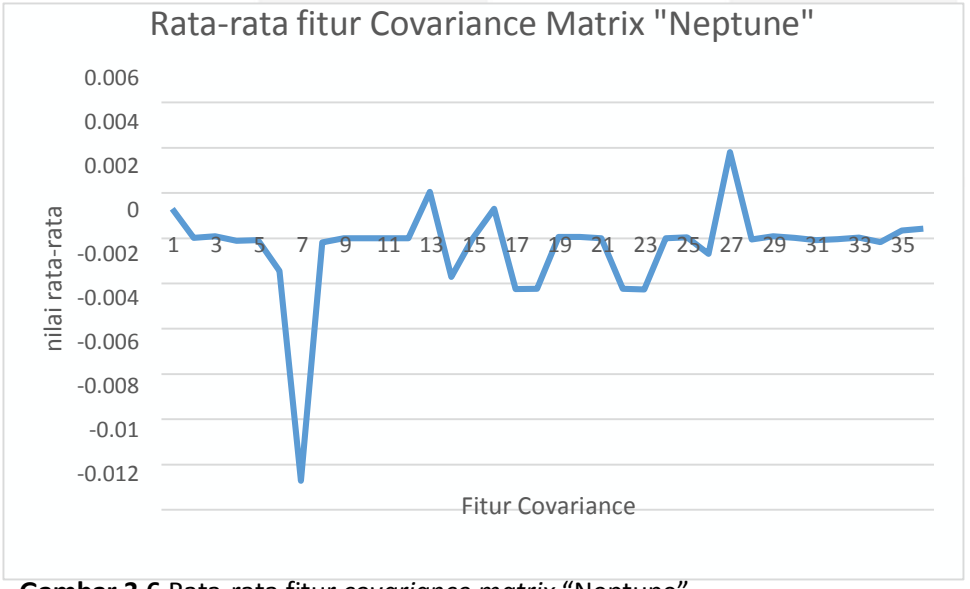
Gambar 3.1 Rata-rata fitur *covariance matrix* "Normal"

Pada Gambar 4.5 ditunjukkan histori 36 fitur trafik *covariance matrix* "Smurf" dari hasil preprocessing *covariance matrix* untuk setiap window yang di proses. Lalu setiap fitur di hitung rata-ratanya pada Gambar 4.6 agar dapat terlihat lonjakan data.



Gambar 3.5 Rata-rata fitur *covariance matrix* "Smurf"

Pada Gambar 4.7 ditunjukkan histori 36 fitur trafik *covariance matrix* "Neptune" dari hasil preprocessing *covariance matrix* untuk setiap window yang di proses. Lalu setiap fitur di hitung rata-ratanya pada Gambar 4.8 agar dapat terlihat lonjakan data.



Gambar 3.6 Rata-rata fitur *covariance matrix* "Neptune"

4. Kesimpulan dan Saran

4.1 Kesimpulan

Kesimpulan yang dapat diambil dai penelitian tugas akhir ini yaitu :

1. *Preprocessing* yang dibangun memiliki peran penting dalam deteksi, metode *covariance matrix* yang menggunakan *landmark window* pada 9 fitur KDDCUP 99 tidak mengabaikkan korelasi antar fitur yang saling berkaitan.
2. Histori 36 fitur *covariance matrix* pada hasil pengolahan di ambil karena karena matriks diagonal yang di tunjukkan merupakan nilai varian dari salah satu fitur yang tidak memiliki informasi mengenai korelasi antar fitur. Sehingga dapat menghasilkan hasil yang optimal untuk ketiga nilai parameter DR, ACC, dan FPR.
3. Performansi sistem deteksi pada data homogen yang menggunakan dataset KDDCUP 99 pada pengujian memiliki nilai *detection rate* yang tinggi untuk setiap data yang diuji.
4. Sistem dikatakan stabil karena proses yang dilakukan sistem tidak dipengaruhi oleh banyaknya data karena menggunakan landmark window.

5. Performansi sistem deteksi pada data heterogen dengan noise 10%, 20%, dan 30% memiliki tingkat FPR yang tinggi karena data yang di uji memiliki label normal. Untuk tingkat detection rate hasil tidak bagus karena data yang diuji pada skenario tersebut hanya berlabel normal yang menyebabkan nilai TP tidak besar.
6. Algoritma SVM digunakan untuk *intrusion detection system*, dalam hal ini digunakan untuk mendeteksi serangan DDoS, dan trafik *flashcrowd*.
7. Hasil pengujian *flashcrowd* membuktikan bahwa data *flashcrowd* bukanlah data serangan dan merupakan data normal dengan tingkat *request* yang tinggi

4.2 Saran

1. Pengujian model terbaik menggunakan dataset lain selain KDDCUP 99 dan WorldCup 98.
2. Membangun sistem deteksi untuk *real time* dan membuat sistem *prevention* (pencegahan) terhadap anomali yang terjadi pada jaringan.

Daftar Pustaka

[1] Yudha Purwanto, Kuspriyanto, Hendrawan, dan Budi Rahardjo, "Traffic Anomaly Detection in DDoS Flooding Attack," *THE 8TH INTERNATIONAL CONFERENCE ON TELECOMMUNICATION SYSTEM,SERVICES,AND APPLICATION*, 2014.

[2] S. Jin, D. S. Yeung and X. Wang, "Network Intrusion Detection in Covariance Feature Space," *Pattern Recognition*, pp. 2185-2197, 2007.

[3] E. Council, "Denial of Service," in *Ethical Hacking and Countermeasures v7.1*, 2011, pp. 433-440.

[4] G. M. A. T. E. A. Sajal Bhatia, "Parametric Differences Between a Real-world Distributed Denial of Service Attack and Flash Event," in *International Conference on Availability, Reliability and Security*, 2011.

[5] M. Arlitt and T. Jin, Workload Characterization of the 1998 World Cup Web Site, France, 1999.

[6] M. Tavallaee, W. Lu, S. A. Iqbal and A. A. Ghorbani, "A Novel Covariance Matrix based Approach for Detecting Network Anomalies," *Communication Networks and Services Research Conference*, p. 1, 2008.

[7] S. Jin, S. D. Yeung and X. Wang, "Network intrusion detection in covariance feature space," *Pattern Recognition*, pp. 2185-2197, 2007.

[8] P. P. Widodo, H. and R. T. Handayanto, Penerapan Data Mining Dengan Matlab, Bandung: Rekayasa Sains, 2013.

[9] Y. Purwanto, Kuspriyanto, Hendrawan and B. Rahardjo, "Survey : Metode dan Kemampuan Sistem Deteksi Anomali Trafik".

[10] K. Garg and R. Chawla, "DETECTION OF DDOS ATTACKS USING DATA MINING," vol. 2, no. 1, 2011.