PENGELOMPOKAN TRAFIK BERDASARKAN WAKTU DENGAN ALGORITMA CLUSTREAM UNTUK DETEKSI ANOMALI PADA ALIRAN TRAFIK

TIME BASED TRAFFIC CLUSTERING USING CLUSTREAM ALGORITHM FOR ANOMALY DETECTION ON STREAMING TRAFFIC

Amalia Rizqi Utami[1], Yudha Purwanto[2], Nurfitri Anbarsanti[3]
[1,2,3]Prodi S1 Sistem Komputer, Fakultas Teknik, Universitas Telkom
[1]amaliaru@students.telkomuniversity.ac.id, [2]omyudha@telkomuniversity.ac.id, anbarsanti@telkomuniversity.ac.id

Abstrak

Pada perkembangan teknologi jaringan internet sekarang ini banyak membahas tentang fenomena-fenomena serangan atapun ancaman terhadap sebuah komputer atau server. Banyak sekali macam-macam tipe ancaman pada komputer dalam sebuah jaringan internet seperti *DoS (Denial of Service)*, *DDoS (Distributed Denial of Service)*, *flash-crowd*, dan sebagainya. Oleh karena itu, untuk memudahkan dalam pengambilan informasi agar sesuai dengan keinginan, perlu adanya pengelompokan dalam anomali trafik tersebut untuk mengenali tipe-tipe serangan yang baru.

Dari permasalahan tersebut perlu suatu sistem deteksi anomali trafik yang mempunyai kemampuan untuk mendeteksi anomali dan mengenali setiap serangan yang datang dengan dilakukan pengelompokkan berdasarkan waktu dan grup. Waktu dan grup adalah parameter untuk meningkatkan akurasi deteksi algoritma. Dan pada penelitian ini dibangun sebuah metode IDS yang menggunakan algoritma clustream.

Hasil dari penelitian ini, sistem yang dibangun secara *real-time* dapat bekerja dengan baik dalam deteksi dan membedakan antara trafik normal dan anomali trafik. Pengelompokkan trafik dilakukan perdua detik, setelah itu akan dianalisis dengan algoritma clustream. Algoritma ini terbagi menjadi *online (micro-clustering)* dan *offline (macro-clustering)*. Di mana *macro-clustering* akan menggunakan data hasil dari *micro-clustering*.

Kata Kunci : anomali trafik, *clustering*, algoritma clustream, *stream traffic*

Abstract

In the development of internet network technology is now widely discusses the phenomena of attacks or threats against a computer or server. There are so many kinds of types of threats on a computer in an Internet network such as DoS (Denial of Service), DDoS (Distributed Denial of Service), flash-crowd, and so forth. Therefore, to facilitate the retrieval of information in order to conform with the desire, the need for the grouping in the traffic anomalies to identify the types of new attacks.

Of these problems need a traffic anomaly detection system that has the ability to detect anomalies and identify any attack that comes with the grouping is done based on time and group. Time and groups are the parameters to improve the accuracy of detection algorithms. And in this study constructed a method that uses an algorithm clustream IDS.

The results of this study, the system is built in real-time can work well in the detection and distinguish between normal traffic and traffic anomalies. Grouping traffic carried perdua seconds, after which it will be analyzed by the algorithm clustream. This algorithm is divided into online (micro-clustering) and offline (macro-clustering). Where macro-clustering will use data from the micro-clustering results.

Keywords: *anomaly traffic, clustering, clustream, stream traffic*

1. Pendahuluan

Keamanan jaringan komputer saat ini mulai banyak diminati oleh banyak kalangan. Jika dilihat dari segi negatifnya, maka akan semakin banyak jenis penyusupan ataupun serangan yang dapat dilakukan dalam suatu jaringan melihat perkembangan teknologi yang semakin canggih dan modern. Untuk itu suatu sistem keamanan jaringan harus dapat melindungi data terhadap serangan atau penyusup di jaringan oleh pihak yang tidak berwenang.

Bentuk serangan dari luar jaringan dapat bersifat merugikan seperti pengambilan data/informasi tanpa ijin. Jenis serangan juga dapat berkembang semakin luas seiring dengan perkembangan teknologi. Contoh jenis serangan yang merugikan pada jaringan antara lain: *DOS attack, CGI attacks, SMB probes, OS fingerprinting*, dll. Serangan-serangan tersebut tentu sangat merugikan jika tidak dapat terdeteksi oleh jaringan komputer yang sedang digunakan. Pihak luar akan mengambil keuntungan tanpa jejak pada suatu jaringan.

Dalam survey [1] sistem deteksi anomali trafik ini dilakukan suatu pendekatan ke masalah deteksi serangan pada jaringan komputer atau dikenal sebagai *Intrusion Detection System* (IDS). Penerapan IDS sebagai *security tools* yang akan mendeteksi instrusi-instrusi, pemindaian, penyerangan ataupun penyusupan serta berbagai ancaman lain pada lalu lintas jaringan seperti anomali trafik. Kekurangan dari penelitian yang sudah ada, deteksi anomali trafik masih dilakukan secara offline. Sehingga dibutuhkan sistem deteksi secara real-time, saat terdapat aktifitas yang mencurigakan pada jaringan akan langsung digenerate oleh sistem.

2. Dasar Teori

2.1 Anomali Trafik

Secara bahasa, anomali trafik merupakan sebuah kejanggalan alur pada trafik data jaringan [12]. Ini disebabkan oleh adanya aktifitas-aktifitas dalam jaringan yang menyimpang dari batas normal. Anomali yang terjadi bisa dilihat melalui kenaikan lonjakan pengguna *internet*, melalui serangan pada suatu trafik dan lonjakan yang tidak disengaja. Kenaikan lonjakan dapat dilihat pada saat adanya bencana yang terjadi kejadian yang tidak biasa terjadi. Kenaikan lonjakan yang terjadi menimbulkan penurunan performansi dari suatu jaringan. Untuk itu perlu dilakukan deteksi terhadap anomali yang terjadi.

2.2 *Preprocessing*

Preprocessing dilakukan untuk mengolah *raw data* menjadi data yang mudah diinterpretasikan sebagai inputan algoritma clustream pada sistem deteksi anomali trafik untuk dianalisis. *Preprocessing* dilakukan terhadap *stream traffic* yang selalu datang. Dengan adanya *preprocessing* akan meningkatkan hasil analisis yang dilakukan. Hasil dari *preprocessing* ini akan digunakan sebagai inputan *clustering* pada algoritma *clustream*.

2.3 Clustering

Clustering adalah suatu teknik analisis dalam pengelompokan objek berdasarkan informasi yang diperoleh. Pada dasarnya objek akan saling berhubungan satu sama lain untuk memaksimalkan dan meminimalkan kesamaan dari anggota *cluster*. *Clustering* dapat dilakukan pada data yang memiliki beberapa atribut yang dipetakan sebagai ruang multidimensi [13]

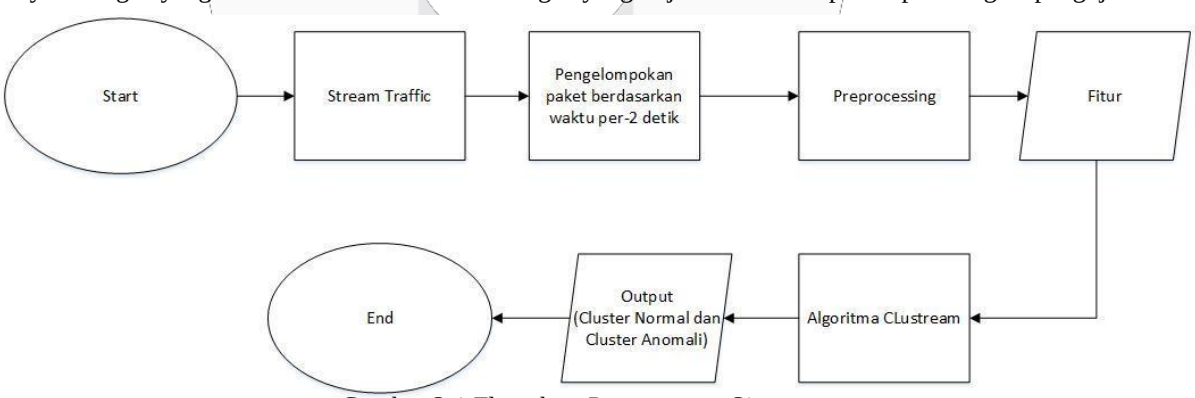
2.4 Algoritma Clustream

Algoritma Clustream digunakan sebagai detektor anomali yang terdiri dari *microclustering (online)* dan *macroclustering (offline)*. Kelompok mikro didefinisikan sebagai perpanjangan sementara *vector* fitur kelompok [9]. Sifat aditivitas dari kelompok mikro tersebut membuatnya menjadi pilihan alami untuk masalah aliran data. Komponen pengelompokan mikro *online* memerlukan proses yang sangat efisien untuk penyimpanan ringkasan statistik yang tepat dalam aliran data yang cepat. Komponen *offline* menggunakan ringkasan statistik ini dalam hubungannya dengan *input* pengguna lain untuk menyediakan pengguna pemahaman yang cepat dari kelompok kapanpun jika diperlukan. Karena komponen *offline* hanya memerlukan ringkasan statistik sebagai *input*, hal tersebut ternyata menjadi sangat efisien dalam praktek. Pendekatan bertahap dua ini juga menyediakan pengguna dengan fleksibilitas untuk mengeksplorasi sifat evolusi dari kelompok selama periode waktu yang berbeda. Ini memberikan wawasan yang cukup untuk pengguna dalam aplikasi nyata.

3. Pembahasan

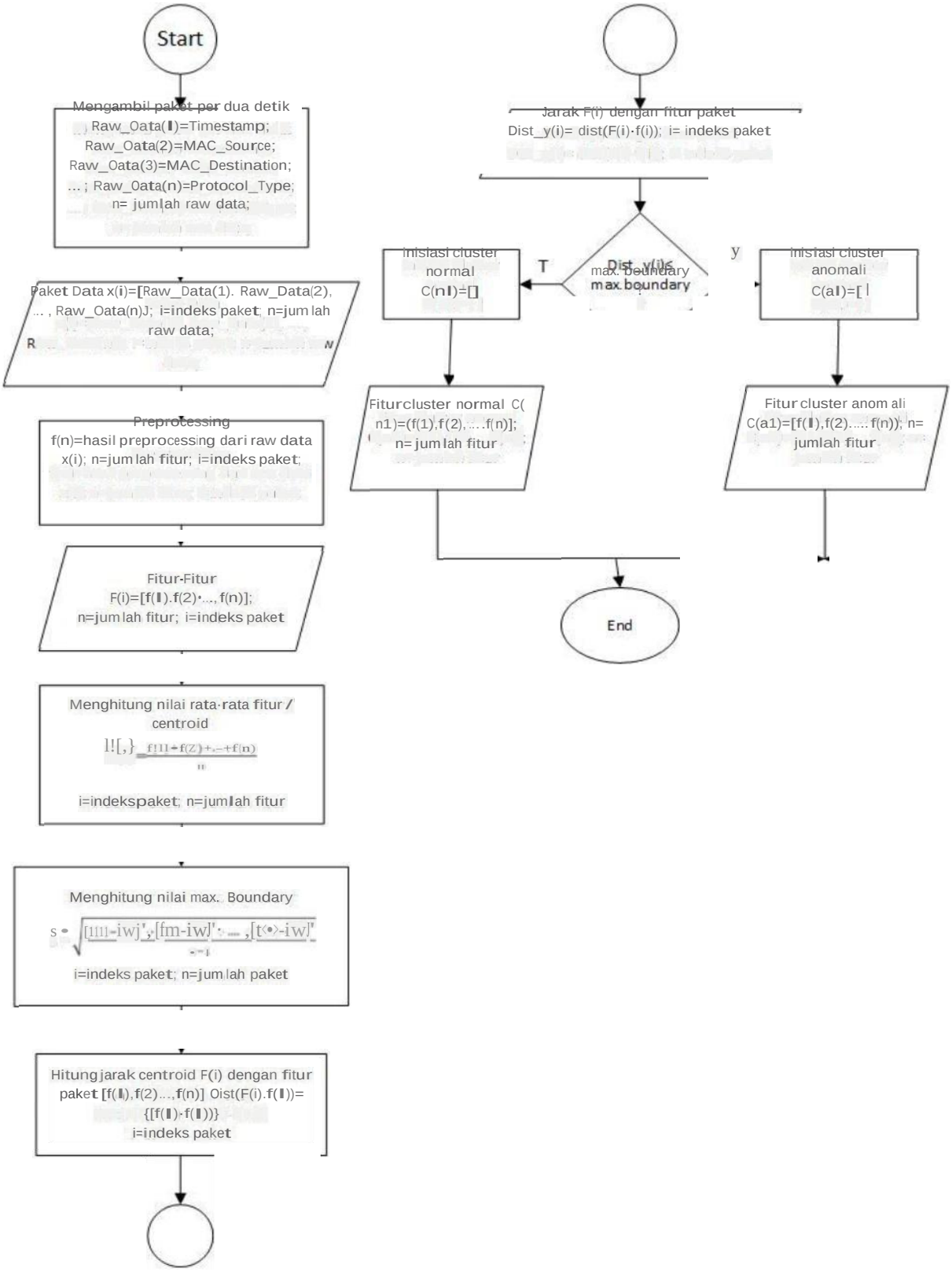
3.1 Deskripsi Sistem

Sistem memiliki tahapan sistem mulai dari men-*generate stream traffic*, melakukan proses normalisasi data, melakukan proses pengklasteran sampai pada proses pengelompokan trafik. Dimana semua proses tadi akan berjalan setelah adanya serangan yang dilakukan oleh attacker. Serangan yang terjadi akan ditampilkan pada bagian pengujian.

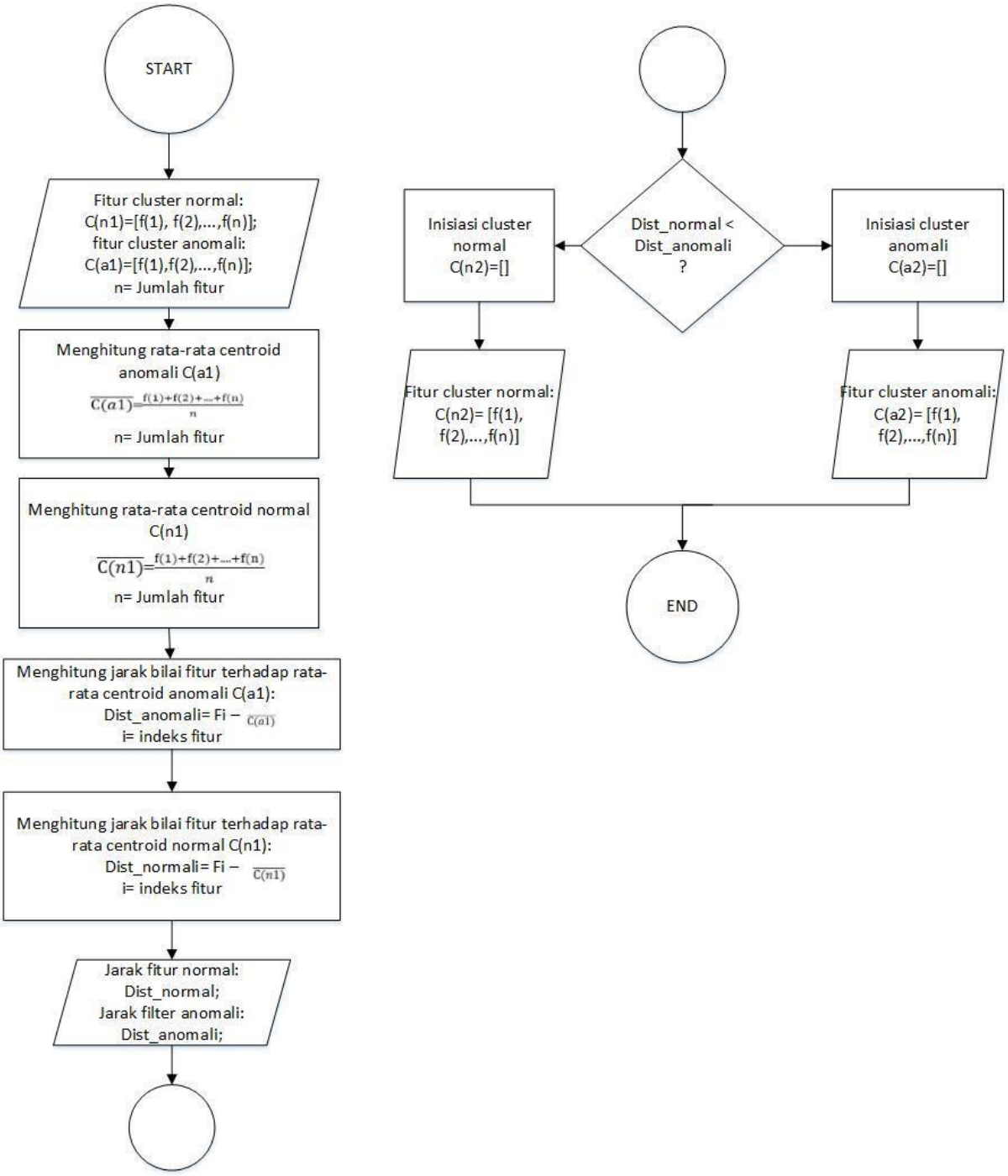


Gambar 3.1 Flowchart Perancangan Sistem

3.2 Flowchart Algoritma Clustream



Gambar 3.2 Flowchart Micro-cluster



Gambar 3.3 Flowchart Macro-cluster

3.3 Confusion Matrix

Dalam penelitian ini, untuk menguji seberapa baik hasil algoritma dalam mendeteksi serangan dibutuhkan metodologi *confusion matrix. Confusion matrix* digunakan untuk menguji tingkat *accuracy* dan *detection rate* pada *cluster*. Berikut adalah rumus-rumus yang digunakan untuk menghitung *accuracy* dan *detection rate.*

$$\text{Akurasi} = \frac{TP+TN}{T\text{[illegible]}+TP+F\text{[illegible]}+FP} \quad (1)$$

$$\textit{Detection Rate} = \frac{TP}{TN+FP} \quad (2)$$

$$\textit{False Positive Rate} = \frac{FP}{T\text{[illegible]}+FP} \quad (3)$$

4. Pengujian dan Analisis

Pada pengujian dibagi menjadi 2 skenario, pengujian dilakukan dengan membandingkan hasil output akhir dengan *dataset* trafik normal hasil dari *preprocessing* secara *online*. Proses *preprocessing* mengkonversikan *raw data* yang di-*capture* oleh sistem secara langsung menjadi fitur-fitur yang mengacu pada *dataset KDDCUP 1999.* Untuk menghasilkan trafik normal, dilakukan *ping IP* biasa terhadap sistem. Proses *preprocessing* akan terus dijalankan hingga mendapatkan *dataset* normal sebanyak 13.700 data paket. Untuk menghasilkan trafik serangan, digunakan ping flood pada terminal oleh user yang bertindak sebagai penyerang. Serangan yang dilakukan pada pengujian memiliki spesifikasi sebagai berikut:

Ping flood dilakukan menggunakan terminal oleh penyerang dengan spesifikasi:

1. *IP source* sudah ditentukan
2. *IP destination* sudah ditentukan
3. Jumlah paket 13.000
4. Besar data 65.000

4.1 Pengujian Trafik Normal

Pada skenario pertama sistem akan mendeteksi trafik normal, yang nantinya akan dibandingkan dengan *dataset* normal sebanyak 13.700 data paket yang sudah diperoleh di awal.

Berikut adalah rincian data yang diperolah untuk dianalisa:

	Aktual	Prediksi
Normal	14775	13205
Anomali	0	1570

Tabel 4.1 Tabel data trafik normal

Aktual	Prediksi	
	Serangan	Normal
Serangan	0	0
Normal	1570	13205

Tabel 4.2 Hasil deteksi 13.755 trafik normal

Detection rate	0,00%
Accuracy	89,37%
False positive rate	10,63%

Tabel 4.3 Nilai *Detection rate, Accuracy* dan *False Positive rate* data trafik normal

Hasil dari *Detection rate* 0,00%, *Accuracy* 89,37% dan *False Positive rate* 10,63% sistem dinilai sangat baik dalam mendeteksi paket normal.

4.2 Pengujian Trafik *Ping Flood*

Pada skenario kedua sistem akan mendeteksi trafik *ping flood*, yang nantinya akan dibandingkan dengan *dataset* normal sebanyak 13.700 data paket yang sudah diperoleh di awal. Berikut adalah rincian data yang diperolah untuk dianalisa:

	Aktual	Prediksi
Normal	0	1771
Anomali	21174	19403

Tabel 4.4 Tabel data trafik *ping flood*

Aktual	Prediksi	
	Serangan	Normal
Serangan	19403	1771
Normal	0	0

Tabel 4.5 Hasil deteksi 13.755 trafik *ping flood*

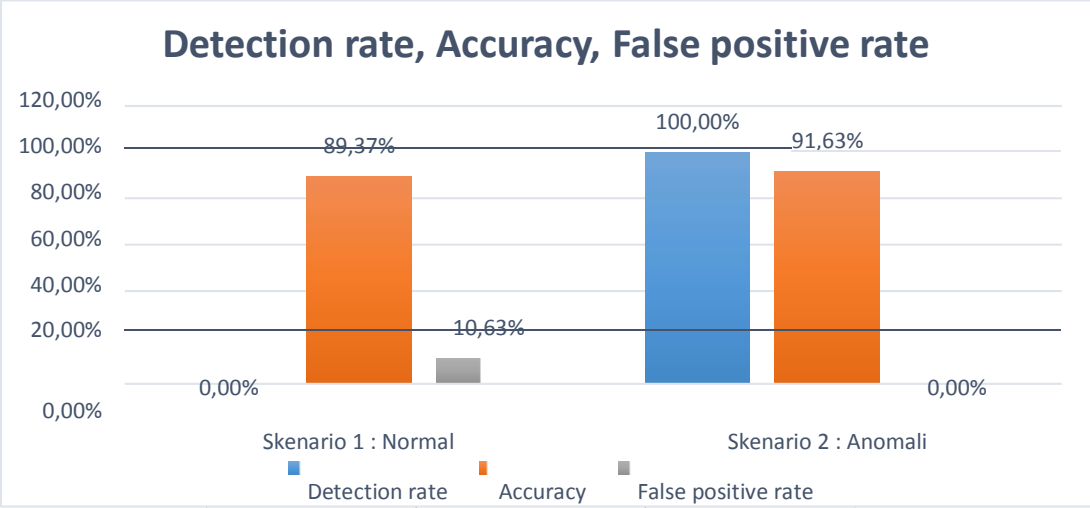
Detection rate	100,00%
Accuracy	91,63%
False positive rate	0,00%

Tabel 4.6 Nilai *Detection rate*, *Accuracy* dan *False Positive rate* data trafik *ping flood*

Hasil dari *Detection rate* 100,00%, *Accuracy* 91,63% dan *False Positive rate* 0,00% sistem dinilai sangat baik dalam mendeteksi paket anomali.

4.3 Analisis

Hasil yang didapat dari pengujian untuk skenario sebelumnya adalah sebagai berikut:



Gambar 5.1 Persentase Detection rate, Accuracy dan False positive rate

Nilai dari persentase dari *Detection rate*, *Accuracy* dan *False positive rate* berubah-ubah sesuai dengan jenis trafik yang digunakan untuk pengujian. Persentase pada *Detection rate* skenario 2 memiliki nilai yang tinggi yaitu 100,00% dan persentase tertinggi pada *Accuracy* ada pada skenario 2 yaitu 91,63%. Nilai *False positive rate* untuk setiap pengujian berada pada rentang nilai (0%-10,63%) yang membuktikan tingkat kesalahan deteksi rendah (<15%). Dari ketiga parameter di atas, sistem dinilai layak untuk mendeteksi paket-paket anomali secara *real-time*. Berdasarkan pengujian yang telah dilakukan pengelompokan berdasarkan waktu per-2 detik mempunyai kekurangan yaitu saat terjadi serangan namun sistem harus memotong paket setelah waktu yang ditentukan dapat mengakibatkan ketidakutuhan paket yang akan diolah selanjutnya. Hal tersebut mengakibatkan nilai akurasi pada pendeteksian serangan menjadi tidak sempurna (≠100%).

5 Kesimpulan dan saran

5.1 Kesimpulan

Dari hasil yang didapatkan pada penelitian ini dapat ditarik kesimpulan sebagai berikut:

Implementasi algoritma *clustream* berdasarkan waktu untuk *clustering* pada *streaming traffic* dinilai efektif karena berhasil mengelompokan trafik anomali dan trafik normal dengan tingkat akurasi tertinggi sebesar 91,63% (pada skenario 2), *detection rate* sebesar 100,00% (pada skenario 2) dan nilai *false positive rate* yang rendah yaitu 0,00% (pada skenario 2).

1.2 Saran

Saran untuk penelitian sebelumnya adalah sebagai berikut:

1. Penelitian lebih lanjut dapat ditambah jumlah serangan yang dilakukan.
2. Penelitian dapat menggunakan algoritma yang berbeda untuk pengimplementaian sistem deteksi.
3. Penggunaan fitur yang tepat pada penelitian selanjutnya untuk meningkatkan tingkat akurasi sistem.
4. Pada penelitian selanjutnya dapat mengklasifikasi setiap jenis anomali.
5. Penggunaan parameter yang berbeda dalam pengelompokan trafik.

DAFTAR PUSTAKA

[1] Yudha Purwanto, Kuspriyanto, Hendrawan, Budi Rahardjo, "Traffic Anomaly Detection in DDos Flooding," *International Conference on Telecommunication Systems Services and Applications (TSSA),* vol. 8, pp. 313-318, 2014.

[2] Barford, P., & Plonka, D., "Characteristics of Network Traffic Flow Anomalies," *ACM SIGCOMM Internet Measurement Workshop,* 2006.

[3] Flickenger, R., & Team, "Wireless Networking in the Developing World Second Edition,"*England: wsfii organization*, 2007.

[4] L. O'Callaghan et al, "Streaming-Data Algorithms For High-Quality Clustering," *ICDE Conference*, 2002.

[5] C. C. Aggarwal, "A Framework for Diagnosing Changes in Evolving Data Streams,"*ACM SIG-MOD Conference*, 2003.

[6] B. Babcock et al, "Models and Issues in Data Stream Systems," *ACM PODS Conference*, 2002.

[7] P. Domingos, G. Hulten, "Mining High-Speed Data Stream," *ACM SIGKDD Conference*, 2000.

[8] S. Guha, N. Mishra, R. Motwani, L. O'Callaghan, "Clustering Data Streams," *IEEE FOCS Conference*, 2000.

[9] T. Zhang, R. Ramakrishnan, M. Livny, "BIRCH:An Efficient Data Clustering Method for Very Large Databases," *ACM SIGMOD Conference*,1996.

[10] Dony Ariyus, "Intrusion Detection System," *ANDI*, Yogyakarta, 2007.

[11] O. Siriporn, and S. Benjawan, "Anomaly Detection and Characterization to Classify Traffic Anomalies Case study: TOT Public Company Limited Network," *World Academy of Science,* Engineering and Technology, 2008.

[12] Burbeck, K., & Nadjem-Tehrani, S. (2005). ADWICE – *Anomaly Detection with Real-time Incremental Clustering.* Paper, Linkopings Universitet, Computer Information and Science.

[13] C. C. Aggarwal. A Survey of Stream Clustering Algorithms: In C.C. Aggrawal, *A Survey of Stream Clustering Algortithm, In "Data Clustering: Algorithms and Applications"* (pp. 230-253). Yorktown Heights, New York: CRC Press.