# PENGELEOMPOKAN TRAFIK BERDASARKAN KELOMPOK DENGAN ALGORITMA CLUSTREAM UNTUK DETEKSI ANOMALI PADA ALIRAN TRAFIK

## *GROUP BASED TRAFFIC CLUSTERING USING CLUSTREAM ALGORITHM FOR ANOMALY DETECTION ON STREAMING TRAFFIC*

**Rahmat Ramadhan[1], Yudha Purwanto[2], Nurfitri Anbarsanti [3]**
[1,2,3]Prodi S1 Sistem Komputer, Fakultas Teknik, Universitas Telkom
[1]rahmatramadhan@students.telkomuniversity.ac.id, [2]omyudha@telkomuniversity.ac.id, [3]anbarsanti@gmail.com

**Abstrak**

Perkembangan jaringan teknologi internet sudah semakin pesat, keamanan jaringan menjadi fokus penting dalam melindungi serangan terhadap suatu data di jaringan. Saat ini begitu banyak jenis penyusupan atau serangan terhadap suatu jaringan komputer. Keamanan jaringan komputer sangatlah penting untuk menjaga integritas data. IDS (Intrusion Detection System) merupakan sistem komputer yang digunakan untuk mengidentifikasi jika terdapat aktifitas yang mencurigakan pada lalu lintas suatu jaringan. Sistem deteksi anomali trafik ini mempunyai kemampuan untuk mendeteksi anomali yang terjadi dan mengenali setiap serangan sehingga dapat dikelompokkan berdasarkan waktu serangan dan kelompok serangan. Waktu serangan dan kelompok serangan adalah parameter untuk meningkatkan akurasi deteksi. Dan pada penelitian ini dibangun sebuah metode IDS yang menggunakan algoritma Clustream. Hasil dari penelitian ini sistem yang dibangun dapat bekerja dengan baik dalam deteksi dan membedakan antara traffic normal dan traffic anomaly. Setiap serangan akan dianalisi dengan algoritma Clustream berdasarkan waktu serangan dan kelompok serangan. Dimana algoritma Clustream terbagi menjadi online (mikro-Clustering) dan offline (makro-Clustering). Pada online komponen menyimpan statistik summary secara periodik tentang stream data sedangkan untuk offline komponen berdasarkan pada statistik summary yang tersimpan.

Kata Kunci : IDS *(Instrusion Detection System)*, *anomaly trafik*, algoritma clustream

**Abstract**

The development of Internet technology network has been growing rapidly, network security becomes an important focus in protecting the attacks on a data network. We have so many kinds of intrusion or attack against a computer network. Network security is extremely important to maintain the integrity of the data. IDS (Intrusion Detection System) is a computer system that is used to identify if there is suspicious activity on the traffic network. This traffic anomaly detection system has the ability to detect anomaly and identify any attacks that can be grouped based on the time of the attack and the raid group. Time attack and raid groups are the parameters to improve detection accuracy. And in this study constructed a method that uses an algorithm Clustream IDS. The results of this research system built to work well in the detection and distinguish between normal traffic and traffic anomaly. Any attack will be analyzed with Clustream algorithm based on the time of attack and group attack. Where Clustream algorithm is divided into online (micro-clustering) and offline (macro-Clustering). In the online component store periodic statistical summary of the data stream while the offline component is based on a statistical summary stored.

Keywords: IDS (Instrusion Detection System), traffic anomaly, the algorithm clustream

## 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. 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.

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. Pada saat ini penanganan yang ada untuk masalah anomaly traffic hanya secara offline atau tidak realtime. Oleh karena itu dibutuhkannya penelitian ini untuk menganalisa adanya anomali trafik pada suatu jaringan secara realtime atau stream.

## 2. Dasar Teori

### 2.1 Deteksi Anomali Trafik

Anomali trafik adalah suatu keadaan yang terjadi pada sebuah lalu lintas jaringan yang menyebabkan kondisi menjadi tidak 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 di dunia, kompetisi atau pertandingan dan kejadian yang tidak biasa terjadi setiap hari. Secara tidak sadar, kondisi kenaikan lonjakan ini memberikan dampak negatif bagi beberapa pihak.Kenaikan lonjakan yang terjadi tersebut menimbulkan penurunan performansi dari suatu jaringan. Untuk itu, perluu dilakukan deteksi terhadap anomali yang terjadi.

### 2.2 Trafik

Trafik dapat didefiniskan sebagai perpindahan informasi dari satu tempat ke tempat lain melalui jaringan telekomunikasi dan streaming adalah teknologi yang memungkinkan suatu file dapat segera dijalankan tanpa harus menunggu selesai didownload dan terus “mengalir” tanpa ada intrupsi. Sehingga streaming trafik adalah perpindahan data atau informasi dari satu tempat ke tempat lain melalui jaringan yang memungkinkan suatu file dapat segera dijalankan tanpa harus menunggu selesai didownload dan terus “mengalir” tanpa ada intrupsi. Pada saat perpindahan data tersebut, data-data yang sedang dipindahakan dapat tercuri atau terserang oleh pihak tertentu. Serangan dapat diklasifikasikan berdasarkan pada tindakan dan tujuan dari penyerangan.

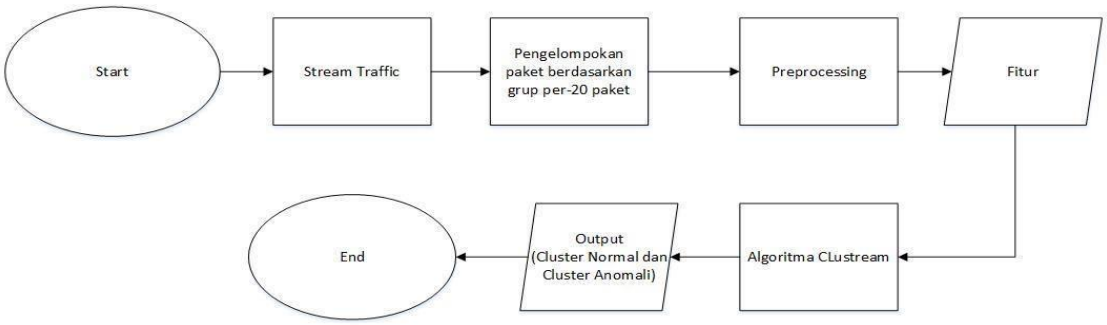
### 2.3 Algoritma Clustream

Algoritma clustream digunakan sebagai detektor anomali yang terdiri dari microclustering (online) dan macroclustering (offline). Kelompok mikro didefinisikan sebagai perpanjangan sementara vektor fitur kelompok . Sifat aditivitas dari kelompok mikro tersebut membuatnya 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 pengguna dengan 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

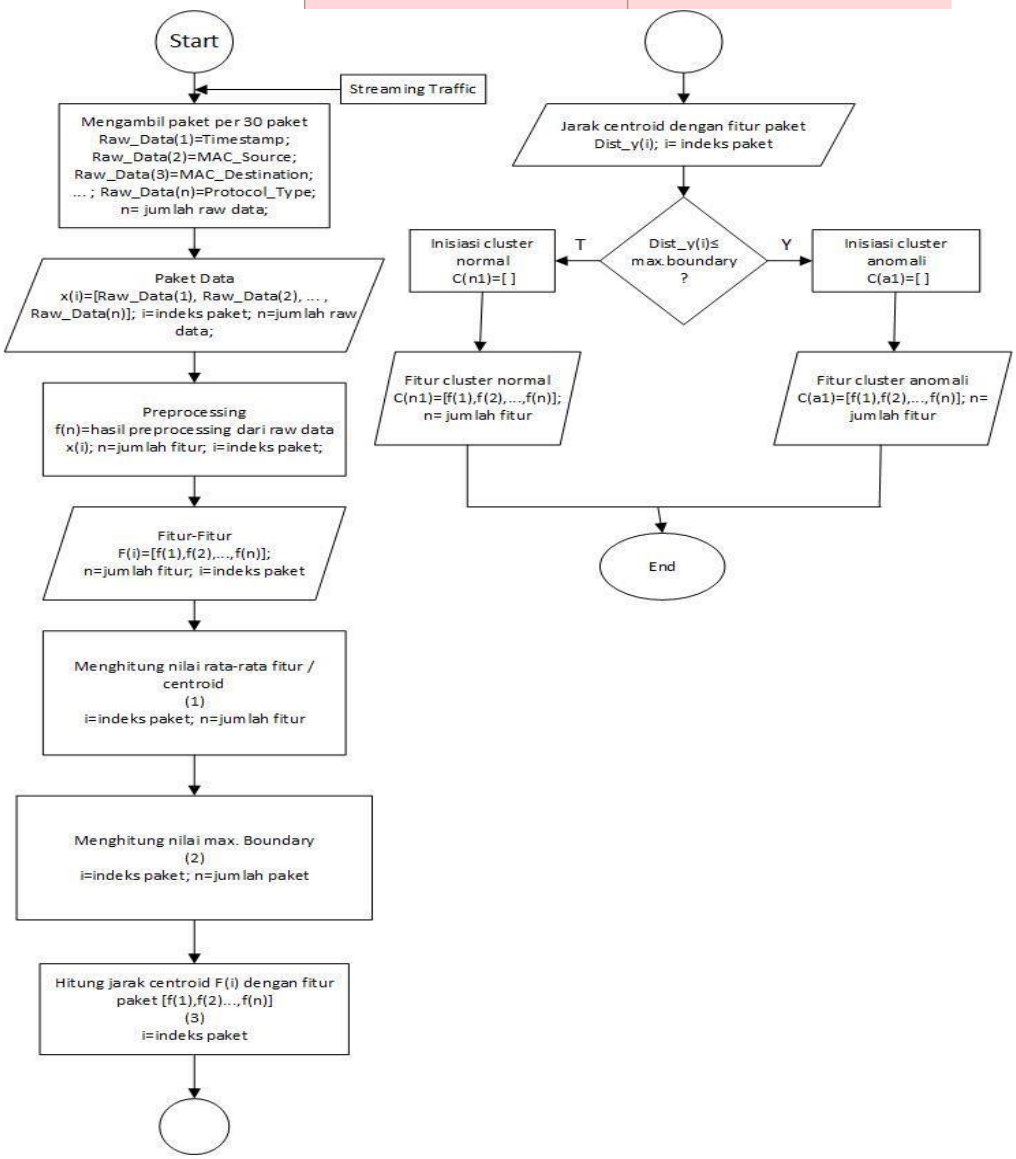
### 3.1 Deskripsi Sistem

Perancangan sistem menggambarkan bagaimana cara kerja generate data serangan dari snort, melakukan proses normalisasi data, melakukan proses pengklasteran sampai pada proses labelisasi. Dimana semua proses tadi akan berjalan setelah adanya serangan yang dilakukan oleh attacker. Serangan yang terjadi akan ditampilkan pada bagian pengujian.

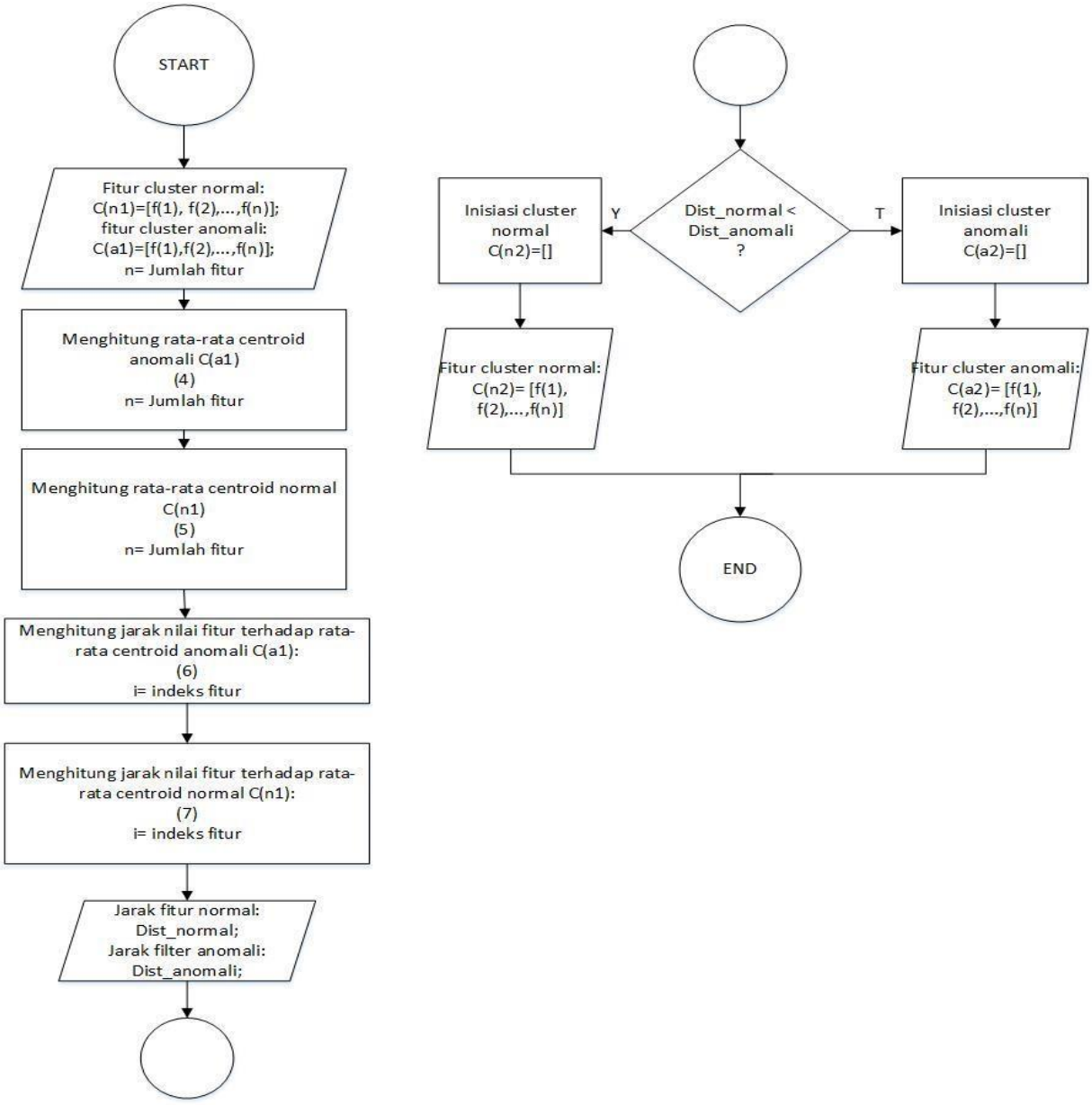



Gambar 3.2 *Flowchart* Sistem

### 3.2 Algoritma Clustream




Gambar 3.2 *Flowchart* Algoritma Clustream (*Microclustered*)




Gambar 3.2 *Flowchart* Algoritma Clustream (*Macroclustered*)

### 3.2 Convetion Matrix

Dalam penelitian ini, untuk menguji seberaoa baik hasil algoritma dalam mendeteksi serangan dibutuhkan metodologi *cofusion matrix*. *Cofusion matrix* digunakan untuk menguji tingkat *accuracy*, *detection rate* serta *false positive rate* pada *cluster*. Berikut adalah rumus-rumus yang digunakan:

$$\text{Akurasi} = \frac{TP + TN}{T\text{□□} + T\text{□□} + F\text{□□} + FN} \quad (1)$$

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

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

## 4. 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:

Tabel 4.1 Data trafik normal per-30 paket

| | **Aktual** | **Prediksi** |
|---|---|---|
| **Normal** | 14400 | 14250 |
| **Anomali** | 0 | 150 |

Tabel 4.2 Hasil deteksi 14400 trafik normal per-30 paket

| **Aktual** | **Prediksi** | |
|---|---|---|
| | **Serangan** | **Normal** |
| **Serangan** | 0 | 0 |
| **Normal** | 150 | 14250 |

Tabel 4.3 Nilai *Detection rate*, *Accuracy* dan *False Positive rate* data trafik normal per-30 paket

| | |
|---|---|
| ***Detection rate*** | 0% |
| ***Accuracy*** | 98.96% |
| ***False Positive Rate*** | 1.04% |

Hasil dari *Detection rate* 0% karena memang tidak terjadi serangan, *Accuracy* 98.96% dan *False Positive rate* 1.04% sistem dinilai sangat baik dalam mendeteksi paket normal.

### 4.2 Pengujian Trafik *Ping Flood*

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:

Tabel 4.1 Data trafik anomali per-30 paket

| | **Aktual** | **Prediksi** |
|---|---|---|
| **Normal** | 0 | 390 |
| **Anomali** | 15870 | 15480 |

Tabel 4.2 Hasil deteksi 14400 trafik anomali per-30 paket

| **Aktual** | **Prediksi** | |
|---|---|---|
| | **Serangan** | **Normal** |
| **Serangan** | 15480 | 390 |
| **Normal** | 0 | 0 |

Tabel 4.3 Nilai *Detection rate, Accuracy* dan *False Positive rate* data trafik anomali per-30 paket

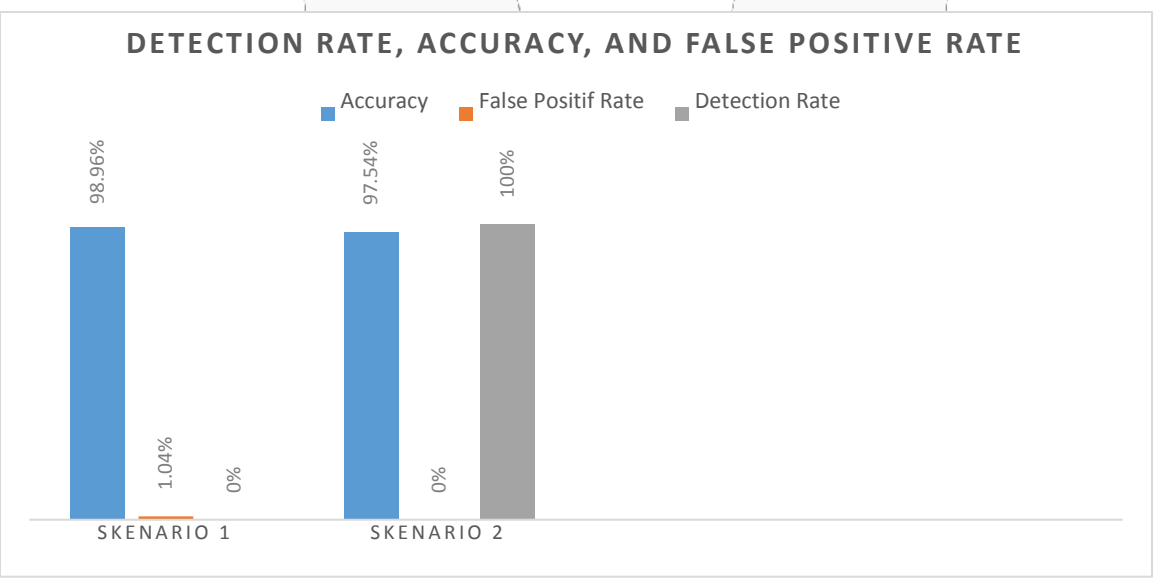
| | |
|---|---|
| ***Detection rate*** | 100% |
| ***Accuracy*** | 97.54% |
| ***False Positive Rate*** | 0% |

Hasil dari *Detection rate* 100%, *Accuracy* 97.54% dan *False Positive rate* 0% sistem dinilai sangat baik dalam mendeteksi paket anomali.

**4.3 Analisis**

Pertama, pada saat pengambilan data normal dan saat terjadi serangan, *input* data yang didapat akan lebih baik karena pengolahan paket perkelompuk menunggu paket yang datang sebanyak 30 paket dengan tidak memperhatikan waktu yang dibutuhkan berlawanan dengan sistem rekan satu tim yang menggunakan waktu 2 detik, dimana akan berpengaruh pada hasil akurasi.

Kedua, hasil yang didapat dari 2 pengujian diatas adalah sebagai berikut:




Gambar 4.18 Persentase *Accuracy*, *False Positive rate*, dan *Detection rate*

Akurasi yang didapat berubah sesuai dengan jenis trafik yang digunakan untuk pengujian. Persentase terbesar ada pada saat pengujian menggunakan trafik normal dengan persentase 98.96% dan persentase terkecil ada pada saat pengujian menggunakan trafik *ping flood* dengan 97.54%. Pada saat trafik serangan saja yang digunakan untuk pengujian, persentase akurasi diatas 90%. Ini menunjukan bahwa sistem dapat mendapatkan akurasi dengan baik jika ada paket normal dan paket anomali yang masuk. Pada nilai persentase *false positive rate* untuk dua pengujian pada didapatkan 0% dan 1.04%. Nilai persentase *detection rate* di dua pengujian adalah 0% dan 100%.
Dari ketiga parameter dan keefektifan pengelompokan paket sebanyak 30 tersebut sistem dinilai layak untuk mendeteksi paket-paket anomali secara langsung atau *real-time*

## 5. Kesimpulan dan saran

### 5.1 Kesimpulan

Dari hasil yang didapatkan pada penelitian ini dapat ditarik beberapa kesimpulan sebagai berikut: Berdasarkan dua pengujian diatas penggunaan algoritma Clustream berdasarkan grup sangat efektif dengan nilai akurasi 97.54%, detection rate di 100%, pada saat sistem diuji dengan ping flood, dan tingkat false positive 1.04%, pada saat sistem diuji dengan ping normal secara real-time. Hal ini menunjukan penggunaan algoritma Clustream berdasarkan grup lebih efektif dibandingkan menggunakan algoritma berdasarkan waktu.

### 5.2 Saran

Saran untuk penelitian selanjutnya adalah :

1. Dapat mengklasifikasikan bentuk serangan yang terjadi.
2. Menambahkan serangan yang dilakukan pada sistem.
3. Dapat menerapkan sistem seperti ini pada *tools* lain yang sejenis.
4. Penggunaan parameter yang berbeda pada penelitian selanjutnya.
5. Penggunaan serangan yang lebih beragam pada penelitian yang selanjutnya.

## DAFTAR PUSTAKA

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

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

[3] Jatmiko Reno Ramadhani. 2015. Analisis Metode Covariance Matrix menggunakan Teknik Landmark Window Untuk Sistem Deteksi Anomali Trafik. Universitas Telkom.

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

[5] M. Agung Tri Laksono. 2015. Analisis Sistem Deteksi Anomali Trafik Menggunakan Algoritma CURE (Clustering Using Representatives) dengan Outlier Removal Clustering dalam Menangani Outlier. Universitas Telkom.

[6] P. Domingos, G. Hulten. Mining High-Speed Data Streams. ACM SIGKDD Conference, 2000.

[7] Putu Ananda Kusuma Wiradharma. 2015. Analisis Sistem Deteksi Anomali Trafik Menggunakan Algoritma Clustering Isodata (Self-Organizing Data Analysis Technique) Dengan Euclidean Distance. Universitas Telkom.

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

[9] 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.

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

[11] 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.

[12] 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.