Analisis Sistem Deteksi Anomali Trafik Menggunakan Pembandingan Perubahan Fitur Pada Triangle-Area-Map Untuk Mengenali Tipe Anomali

Analysis System Anomaly Traffic Detection with Comparing The Diffrences of Triangle-Area-Map Features for Anomaly Type Identification

Mujp Muhammad Irsyad[1], Yudha Purwanto[2], Tito Waluyo Purboyo[3]

[1,2,3]Prodi S1 Teknik Komputer, Fakultas Teknik Elektro, Universitas Telkom
Jl. Telekomunikasi, Dayeuh Kolot, Bandung, 40257 Indonesia

[1]mujpmirsyad@student.telkomuniversity.ac.id, [2]omyudha@telkomuniversity.ac.id,
[3]titowaluyo@telkomuniversity.ac.id

Abstrak

Seiring berkembangnya teknologi internet sekarang ini, semakin banyak muncul jenis serangan ataupun ancaman terhadap sebuah komputer atau server dalam sebuah jaringan, salah satu contohnya berupa anomaly traffic. Beberapa macam tipe anomaly traffic pada komputer dalam sebuah jaringan internet seperti Denial of Service (DoS), Distributed Denial of Service (DDoS) atau flashcrowd. Oleh karena itu, dibutuhkan adanya suatu sistem deteksi untuk mendeteksi dan mengenali setiap anomaly trafik tersebut.

Pada penelitian Tugas Akhir ini mengembangkan sistem deteksi berbasis statistik menggunakan Multivariate Correlative Analysis (MCA). MCA menggunakan teknik representasi Triangel-Area-Map (TAM) untuk mendeskripsikan hubungan antar setiap fitur trafik dengan menghitung jarak satu buah nilai fitur terhadap nilai fitur yang lain untuk setiap fitur hasil ekstraksi. Data hasil pengolahan MCA dianalisis menggunakan Mahalanobis Distance untuk digunakan sebagai data referensi atau observasi. Proses pendeteksian terhadap data yang diobservasi berbasis threshold dari data referensi dan proses klasifikasi anomali menggunakan Mahalanobis Distance dan Cosine Distance untuk menghitung besar jarak antara nilai fitur TAM trafik yang diobservasi dengan TAM trafik acuan. Pengujian sistem dilakukan dengan melakukan pengukuran tingkat keakuratan algoritma, berdasarkan hasil keluaran sistem dengan parameter Detection Rate (DR), False Positive Rate (FPR) dan Accuracy (ACC).

Kata Kunci : anomali trafik, DDoS, flash-crowd, multivariate correlative analysis, triangle-area-map, mahalanobis distance

Abstract

As the development of Internet technology today, a growing number of emerging types of attacks or threats against a computer or a server in a network, one example in the form of traffic anomalies. Some types of anomaly traffic on a computer in an internet network such as Denial of Service (DoS), Distributed Denial of Service (DDoS) or flashcrowd. Therefore, it is necessary to have a detection system to detect and recognize each traffic anomalies.

At this final project research develop statistics-based detection system using Multivariate Correlative Analysis (MCA). MCA using representation techniques Triangle-Area-Map (TAM) to describe the relationship between each feature traffic by calculating the distance of a single feature feature value to the value of other features for each feature extraction results. The results of data processing were analyzed using the Mahalanobis Distance to be used as reference or observation data.

The detection process of the observed data based on threshold of the reference data and anomaly classification process using Mahalanobis Distance and Cosine Distance to calculate the distance between values of the TAM observed traffic features with the TAM reference traffic. System testing is made by measuring the level of accuracy of the algorithm, based on the output of system with parameters Detection Rate (DR), False Positive Rate (FPR) and Accuracy (ACC).

Keywords: traffic anomalies, DDoS, flash-crowd, correlative multivariate analysis, triangle-area-map, mahalanobis distance

1. Pendahuluan

Perkembangan internet saat ini sebagai suatu media informasi sangatlah pesat. Setiap orang dapat memanfaatkannya untuk berbagai kepentingan atau aspek kehidupan, sehingga kian hari trafik internet menjadi semakin tinggi. Intensitas trafik dapat berubah menjadi jauh lebih tinggi daripada keadaan normal, biasanya terjadi karena dua hal yaitu fenomena flashcrowd atau serangan flooding trafik.

Flashcrowd adalah kejadian yang tidak dapat diprediksi tetapi sah/legal yaitu peningkatan akses oleh user yang sah/legal secara dramatis/tinggi ke suatu server karena suatu kejadian seperti bencana alam, peluncuran produk, breaking news, dll. [1], sedangkan salah satu serangan flooding trafik yang menyebabkan kerusakan paling parah adalah *Denial of Service* (DOS) dan *Disributed Denial of Service* (DDoS) yang merupakan suatu bentuk serangan flooding sistematis yang berusaha menghabiskan sumberdaya suatu host atau service dan membuatnya menjadi tak dapat diakses oleh user yang berhak (intended/legitimate user) yang biasanya dilakukan suatu pihak untuk merugikan pihak lain.

Oleh karena itu dibutuhkan suatu sistem atau mekanisme deteksi anomali, salah satunya dengan pendekatan statistik. Pada Tugas Akhir ini sistem deteksi anomali menggunakan pendekatan statistik *Multivariate Correlation Analysis* (MCA) berbasis Triangle-Area-Map (TAM). Memanfaatkan *Mahalanobis Distance* untuk menghitung perbedaan hasil proses MCA antar tiap trafik. Proses deteksi yang dilakukan berdasarkan *threshold* trafik normal acuan dan proses klasifikasi anomali berdasarkan jarak TAM trafik anomali terhadap TAM trafik normal menggunakan *Mahalanobis Distance* dan *Cosine Distance*. Kemudian kemampuan sistem dianalisis dari hasil keluaran sistem dengan parameter *Detection Rate*, *False Positive Rate* dan *Accuracy*.

2. Dasar Teori dan Perancangan

2.1. Sistem Deteksi Anomali

Dalam pendeteksian serangan atau anomali yang terjadi di dalam jaringan, dikenal dua istilah pendekatan yaitu *Intrusion Detection System* (IDS) dan *Intrusion Prevention System* (IPS) [1]. Keduanya merupakan suatu sistem yang bekerja mengawasi keadaan jaringan dan memberikan alarm kepada sistem administrator apabila sistem mendeteksi anomali atau serangan. Pada IDS penindaklanjutan setelah deteksi dilakukan secara manual oleh sistem administrator sedangkan pada IPS dilakukan secara otomatis oleh sistem. Pada IDS/IPS terdapat dua metode dalam mendeteksi anomali/serangan yaitu *intrusion signature*/*misuse based* dan *traffic anomaly based* [2]. Pada *signature based*, sistem mendeteksi serangan dengan mencocokan pola trafik yang dideteksi/diawasi dengan pola trafik di database anomali/serangan yang dimiliki sistem sedangkan pada *anomaly based*, sistem melakukan pengawasan atau observasi pada jaringan untuk mendeteksi pola anomali yang terjadi dengan pendekatan tertentu seperti statistik. Hal tersebut membuat *anomaly based* sistem memiliki kemungkinan false detection yang tinggi, namun dapat mendeteksi serangan baru.

2.2. Distributed-Denial of Service (DDoS)

DDoS merupakan serangan flooding trafik [3] yang dilakukan dengan sengaja untuk mengganggu QoS dari sistem dengan sasaran link/bandwidth untuk membuat sumber daya server habis bahkan hingga membuat server down/crash serta membuat user yang berhak tidak mendapatkan layanan dari server secara terdistribusi oleh banyak zombie yang dikendalikan oleh botnet, dimana botnet sendiri dikendalikan oleh botmaster dengan beberapa hal seperti membanjiri lalu lintas yang banyak atau memenuhi link dengan paket yang besar [1]. Selain itu juga menurunkan sisi availability servis server yang membuat user tidak dapat menggunakan layanan server tersebut. Hal tersebut yang membuat serangan DDoS lebih berbahaya daripada serangan DoS karena DoS berasal dari satu penyerang sedangkan DDoS berasal dari banyak penyerang yang terkoordinasi bersama-sama, sehingga serangan yang dilakukan lebih sulit ditelusuri keberadaan penyerang yang sebenarnya, terlebih lagi apabila pola serangan dilakukan secara dinamis.

2.3. Flashcrowd

Flashcrowd merupakan suatu keadaan dimana terjadi lonjakan akses yang tiba-tiba secara alamiah, tidak disengaja dan tidak terduga ke suatu server yang datang dari berbagai user yang sah dan berhak yang tersebar tidak terbatas secara acak di Internet pada suatu periode waktu tertentu secara gradual dalam hitungan menit atau jam dan jarang sekali terjadi dalam hitungan detik. Hal ini menyebabkan peningkatan dramatis beban server dan tekanan berat pada link jaringan yang mengarah ke server, sehingga menghasilkan peningkatan substansial dalam packet loss dan kepadatan trafik [4].

2.4. Mahalanobis Distance

Mahalanobis Distance (MD) dapat mengukur jarak diantara dua objek data multivariat atau antara suatu titik A dengan suatu distribusi B sehingga dapat mengetahui korelasi antara dua variabel dan menghapus

ketergantungan pada skala pengukuran selama penghitungan [5] [6]. Dapat juga menghitung seberapa jauh standar deviasi A terhadap rata-rata B atau seberapa besar kesamaan antara sekumpulan kondisi terhadap sekumpulan kondisi yang ideal dengan memikirkan korelasi antar objek dalam bentuk variabel vektor dan matrik covariance dari kedua objek tersebut, seperti diformulasikan sebagai berikut.

$$\text{[illegible]}(\text{[illegible]}) = \sqrt{\frac{(\text{[illegible]}-\mu\text{[illegible]})^{T}(\text{[illegible]}-\mu\text{[illegible]})}{\text{[illegible]}}} \qquad (1)$$

2.5. Cosine Distance

Cosine Distance (CD) merupakan salah satu teknik untuk menghitung jarak antar dua vektor yang pada dasarnya berasal dari sisa nilai dalam skala 1 sampai 0 dari Cosine Similarity. Cosine Similarity sendiri mengukur kesamaan atau kemiripan antara dua vektor dot product dengan menghitung sudut kosinus diantara kedua vektor tersebut. Biasanya Cosine Similarity (3) digunakan dalam ruang positif dengan hasil berada diantara 0 dengan 1 berikut merupakan formula Cosine Distance (2).

$$\text{[illegible]}(\text{[illegible]}) = 1 - \text{[illegible]} \qquad (2)$$

$$\text{[illegible]} = 1 - \frac{\text{[illegible]}}{\|\text{[illegible]}\|\|\text{[illegible]}\|} = 1 - \frac{\sum_{i=1}^{n} \text{[illegible]}}{\sqrt{\sum_{i=1}^{n}(\text{[illegible]})^2} \times \sqrt{\sum_{i=1}^{n}(\text{[illegible]})^2}} \qquad (3)$$

2.6. Dataset KDD Cup 1999 dan World Cup 1998

KDD Cup 1999 merupakan dataset yang banyak digunakan dalam penelitian *Intrussion Detection System*. KDD Cup pada awalnya digunakan pada *The Third Internasional Knowledge Discovery and Data Mining Tools Competition*, yang berlangsung bersamaan dengan *KDD 99 The Fifth Internasional Conference on Knowledge Discovery and Data Mining*. Database ini berisi satu set standar data yang diaudit, yang mencakup berbagai gangguan yang disimulasikan dalam lingkungan jaringan militer [7]. Dataset ini memiliki berbagai macam trafik, mulai dari normal, DDoS, R2L, U2R dan probing yang terdiri dari total 41 fitur [8].

World Cup 1998 Dataset merupakan dataset yang berisi request kepada server 1998 World Cup Web pada saat Piala Dunia 1998 yang di-capture mulai tanggal 30 April sampai 26 Juli 1998 atau sekitar 88 hari dengan total request sekitar 1.352.804.107 request [9]. Dataset ini aslinya berupa Common Log Format dengan jangka waktu perekaman dalam satu detik yang berisi timestamp, clientID, objectID, size, method, status, type, server. Dataset ini banyak digunakan untuk penelitian kejadian flashcrowd karena memiliki data trafik flashcrowd yang terjadi ketika banyak user melakukan request ke 1998 World Cup Web untuk melihat hasil pertandingan.

2.7. Preprocessing

Pada tahap *preprocessing* dataset KDD Cup 1999 dan World Cup 1998, diolah agar dapat dijadikan masukan berupa data testing dan data latih yang berisi kolom-kolom fitur trafik yang dianalisis proses normalilsasi. Data latih dari dataset KDD Cup 1999 berisi seluruh data trafik berlabel normal untuk dijadikan acuan proses deteksi *anomaly traffic* sedangkan data testingnya berisi data trafik berlabel normal, DDoS yaitu Back, Neptune, dan Smurf atau campuran dari keduanya.

Pada *preprocessing* dataset World Cup 1998, data trafik yang berupa *binary log files* di ekstrak ulang menjadi *Common Log Format* dengan program *WorldCup_tools* yang telah tersedia [9]. Lalu dari data trafik *Common Log Format* dilakukan ekstraksi fitur untuk dijadikan data latih yang berisi data trafik normal dan data testing yang berisi trafik normal dan *flashcrowd*. Keadaan *flashcrowd* dapat diketahui dari hasil hitung flow dan paket per detik saat ekstraksi fitur.

Fitur yang digunakan dalam penelitian ini adalah *count, serror_rate, rerror_rate, same_srv_rate, diff_srv_rate, srv_count, srv_serror_rate, srv_rerror_rate,* dan *srv_diff_host_rate* [7].

2.8. Normalisasi

Normalisasi dilakukan karena banyaknya data bias [2] dan tidak berada dalam skala nilai yang sama atau bernilai nol pada data latih dan data tes membuat banyak kekeliruan penghitungan jarak antar nilai fitur, membuat deteksi menjadi kurang efektif, terutama pada beberapa jenis anomali yang tidak menunjukan perubahan nilai yang besar pada fitur-fiturnya. *Statistical normalization* dilakukan untuk mengubah data nilai tiap fitur x data menjadi terdistribusi normal yang standar dengan memanfaatkan nilai rata-rata dan standar deviasi fitur [10]. Normalisasi

membuat 99,9% nilai fitur berada dalam skala [-3,3]. Nilai hasil normalisasi dengan formula rata-rata: $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$ dan formula (4) untuk standar deviasi diperoleh berdasarkan formula (5) berikut.

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)} \quad (4)$$

$$z = \frac{v_i - \mu}{\sigma} \quad (5)$$

2.9. Algoritma MCA berbasis TAM

Multivariate Correlation Analysis (MCA) adalah suatu bentuk dari statistik untuk mengamati dan menganalisis lebih dari satu variabel keluaran atau banyak variabel. MCA dapat menghitung kekuatan hubungan antar tiap variabel dalam suatu kelompok variabel multidimensi. Salah satu kelebihan MCA dalam dalam menganalisis suatu trafik atau fitur-fiturnya adalah tidak perlu catatan sebelumnya atau data *history* dari trafik tersebut saat melakukan analisis trafik serta tahan terhadap perubahan linier semua fitur. MCA juga memberikan karakterisasi catatan trafik jaringan secara individual daripada model perilaku suatu grup catatan trafik [2]. Tidak seperti pendekatan *Covariance Matrix* [11] yang rentan terhadap perubahan linier semua fitur, MCA tahan terhadap masalah ini dengan memanfaatkan TAM yang pada dasarnya berawal dari *Euclidean Distance Map* [12] [13]. Dengan begitu akan mengurangi *latency* pengambilan keputusan juga biaya komputasi saat mendeteksi serangan serta memungkinkan dilakukannya deteksi *sample-by-sample*.

Triangle-Area-Map merupakan teknik yang digunakan untuk merepresentasikan hasil ekstraksi nilai geometris dalam koordinat kartesian dari setiap kemungkinan permutasi penghitungan antara dua fitur trafik dari setiap dataset latih dan test yang telah dinormalisasi.

Algoritma 2 : *Triangle-Area-Map* (*dataset*,*t*)

```
1: Masukkan dataset
2:   repeat
3:     proses j fitur dari x data per t
4:     j = k
5:     for 1 to j do
6:       for 1 to k do
7:         if j = k then
8:           m(j,k) = 0
9:         else if  k < j then
10:          m(j,k) = (fitur j * fitur k)/2
11:          Simpan m(j,k) di matriks TAM
12:        end if
13:      end for
14:    end for
15:    Simpan TAM di t TAMLower
16: until semua data fitur diproses
17: Simpan TAMLower
```

2.10. Sistem Deteksi Dan Klasifikasi

Formula (6) merupakan formula *threshold* yang pada dasarnya berasal dari aturan empiris atau *three-sigma rule*, dengan alpha (α) adalah parameter untuk menentukan rentang penerimaan data yang diobservasi sebagai normal yang biasanya berada dalam rentang 1 sampai 3 kali standar deviasi dari rata-rata, μ adalah rata-rata dan σ adalah standar deviasi dari MD trafik normal. Apabila MD dari trafik yang diobservasi dengan profil trafik normal lebih besar daripada *threshold*, maka biasa dianggap terjadi anomali, baik serangan atau flashcrowd.

$$\text{threshold} = \mu \pm \sigma * \alpha \quad (6)$$

Setelah tahap deteksi dilanjutkan dengan tahap klasifikasi. Klasifikasi trafik ditentukan berdasarkan hasil penghitungan jarak terkecil atau yang paling dekat dari hasil penghitungan jarak antara vektor *TAMlower* tiap trafik yang diobservasi dengan rata-rata tiap fitur *TAMlower* dari tiap profil trafik. Penghitungan jarak pada tahap klasifikasi penelitian ini menggunakan masing-masing *Mahalanobis Distance* dan *Cosine Distance*.

Algoritma 3 : *Klasifikasi Trafik* ($TAM_{Observasi}$, TAM_{Profil}, t)

```
1: Masukkan TAM_Observasi, TAM_Profil
2:   repeat
3:     proses x TAM_Observasi per t
4:     for 1 to x do
5:       for 1 to y TAM_Profil do
6:         DistMD = Hitung MD(TAM_Observasi, TAM_Profil)
7:         Simpan DistMD di matriks DMD
8:         DistCD = Hitung CD(TAM_Observasi, TAM_Profil)
9:         Simpan DistCD di matriks DCD
10:      end for
11:      Cari index DistMD dan DistCD terkecil
12:      if Index DistMD = Index TAM_Profil ke y then
13:        LabelMD = TAM_Profil ke y
14:      end if
15:      if Index DistCD = Index TAM_Profil ke y then
16:        LabelCD = TAM_Profil ke y
17:      end if
18:      Simpan LabelMD,LabelCD di LabelList
19:  until.semua TAM_Observasi diproses
20:  Simpan LabelList
```

3. Pembahasan

Pada penelitian [2] proses MCA dilakukan untuk memperoleh informasi hubungan antar fitur setiap data dan direpresentasikan dengan teknik TAM. Data hasil pengolahan MCA dianalisis menggunakan *Mahalanobis Distance* untuk digunakan sebagai data referensi atau observasi. Deteksi anomali yang dilakukan berbasis threshold dengan acuan data trafik normal. Sedangkan dalam penelitian ini, setelah melakukan proses deteksi pada trafik, dilanjutkan dengan proses klasifikasi berbasis jarak pada setiap data trafik yang diobservasi. Pengujian menggunakan data latih dan data tes dari dataset KDD Cup 1999 dan World Cup 1998.

3.1. Parameter Pengujian

Parameter yang digunakan adalah *Detection Rate* (DR), *False Positive Rate* (FPR) dan *Accuracy* (ACC) yang diperoleh berdasarkan *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

Tabel 3.1.1 Contingency Table

		Actual Value	
		Positive	**Negative**
Prediction Value	**Positive**	TP	FP
	Negative	FN	TN

TP : data trafik anomali dideteksi sebagai trafik anomali. FP : data trafik normal dideteksi sebagai trafik anomali. FN : data trafik anomali dideteksi sebagai trafik normal. TN : data trafik normal dideteksi sebagai trafik normal.

Detection Rate (DR) merupakan persentase jumlah anomali yang dideteksi secara tepat dibandingkan dengan total trafik yang dideteksi sebagai anomali. Detection Rate diformulasikan sebagai berikut.

$$\text{Detection Rate} = \frac{TP}{TP+FP} \quad (7)$$

False Positive Rate (FPR) merupakan persentase banyaknya kesalahan deteksi trafik normal sebagai anomali berbanding dengan total trafik normal. False Positive Rate diformulasikan sebagai berikut.

$$\text{False Positive Rate} = \frac{FP}{FP+TN} \quad (8)$$

Accuracy (ACC) merupakan persentase akurasi deteksi sistem terhadap trafik normal dan anomali dibanding total data hasil deteksi. Accuracy diformulasikan sebagai berikut.

$$\text{[illegible]} = \frac{TP + T\text{[illegible]}}{T\text{[illegible]} + T\text{[illegible]} + F\text{[illegible]} + F\text{[illegible]}} \quad (9)$$

3.2. Pengujian Sistem Deteksi dan Klasifikasi

3.2.1. Serangan DDoS

Data latih campuran normal dan serangan yang terdiri dari Back, Neptune dan Smurf dengan total 487.472 data. Sebanyak 97.278 data berlabel normal digunakan sebagai acuan trafik normal.

Tabel 3.2.1.1 Data Hasil Deteksi Data Latih Campuran Normal dan Serangan

	Alpha				
	1,0	**1,5**	**2,0**	**2,5**	**3,0**
TP	388523	388439	388333	388104	387855
TN	91136	94314	95889	96152	96433
FP	6142	2964	1389	1126	845
FN	1671	1755	1861	2090	2339
DR	98.4437%	99.2427%	99.6436%	99.7107%	99.7826%
FPR	6.3139%	3.0469%	1.4279%	1.1575%	0.8686%
ACC	98.3972%	99.0319%	99.3333%	99.3403%	99.3468%

Dari tabel 3.2.1.1 dapat dilihat berdasarkan perubahan nilai alpha mulai dari 1 dan naik terus sebesar 0,5 hingga 3 terdapat kenaikan nilai TP, TN dan FN sedangkan nilai FP cenderung menurun hal ini menyebabkan persentase DR pada nilai alpha 1 sebesar 98.4437% dan naik terus hingga 99.7826% pada nilai alpha 3.

Kenaikan juga terjadi pada akurasi deteksi, sebesar 98.3972% pada nilai alpha 1 dan naik hingga 99.3468% pada nilai alpha 3. Namun persentase FPR justru menurun dari 6.3139% saat nilai alpha 1 hingga 0.8686% saat nilai alphanya 3. Hal ini menunjukan perubahan nilai alpha pada threshold sangat mempengaruhi kemampuan deteksi sistem. Hal tersebut terlihat pada perubahan nilai TP,TN, FP dan FN yang menyebabkan persentase DR dan ACC meningkat sedangkan FPR menurun yang berarti ketepatan deteksi semakin meningkat dengan kesalahan deteksi yang semakin mengecil. Namun hal tersebut tidak memberi jaminan semakin besar nilai alpha dapat memberi hasil deteksi yang lebih baik, justru hal ini dapat mengakibatkan kesalahan pada deteksi serangan yang memiliki perilaku yang mirip dengan trafik normal karena threshold yang digunakan terlalu lebar.

Tabel 3.2.1.2 Data hasil klasifikasi KDD Cup ’99 Normal dan Serangan

	MD	CD
DR	99.1918%	84.5443%
FPR	3.2556%	73.3115%
ACC	99.0455%	85.3522%

Dapat dilihat pada tabel 3.2.1.2 apabila dibandingkan, persentase DR dan ACC *Mahalanobis Distance* memiliki hasil yang lebih baik daripada *Cosine Distance*. Persentase FPR MD jauh lebih kecil daripada CD meskipun tidak begitu baik yaitu sebesar 3.2556% yang dikarenakan cukup banyaknya data trafik serangan Back yang diklasifikasikan sebagai normal dan beberapa data trafik lainnya seperti yang ditunjukan pada tabel 3.2.1.3.

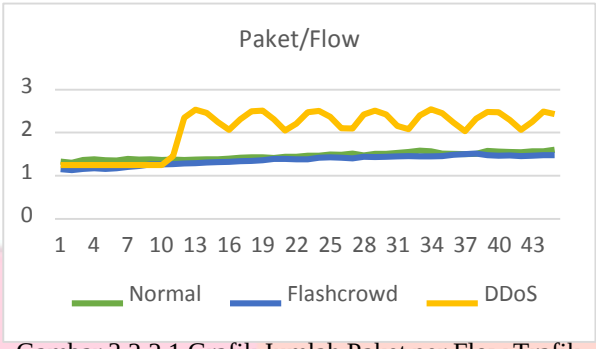
Tabel 3.2.1.3 Data kesalahan deteksi *Mahalanobis Distance*

		Klasifikasi			
		Normal	**Back**	**Neptune**	**Smurf**
Actual	**Normal**	-	0.6292%	0.0064%	0.0142%
	Back	0.2484%	-	0.0000%	0.0000%
	Neptune	0.0016%	0.0000%	-	0.0002%
	Smurf	0.0548%	0.0002%	0.0000%	-

Besarnya persentase FPR CD disebabkan kesalahan klasifikasi yang sangat banyak terutama kesalahan klasifikasi trafik normal yang dianggap sebagai serangan. Hal ini dikarenakan CD kurang baik untuk pengukuran karena tidak memenuhi sifat pertidaksamaan segitiga dan menyalahi *coincidence axiom* dan umum digunakan pada ruang positif saja, meskipun kedua distance tersebut dapat digunakan untuk penghitungan multi-dimesional.

Hal ini menandakan bahwa MD dapat menghitung lebih baik jarak antara vektor data yang diobservasi dengan vektor rata-rata profil normal sehingga dapat menklasifikasikan serangan dengan lebih tepat.

3.2.2. Flashcrowd

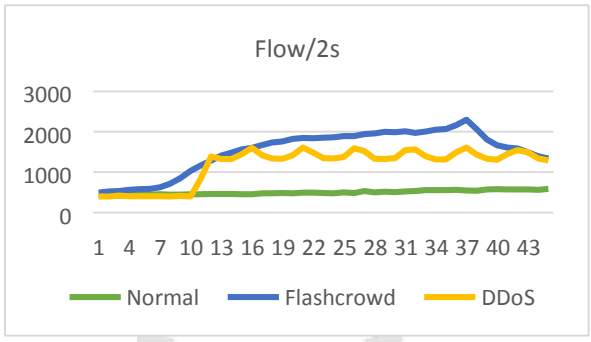


Gambar 3.2.2.1 Grafik Jumlah Paket per Flow Trafik

Proses deteksi dan klasifikas *flashcrowd* yang menggunakan *Mahalanobis Distance* sebagai pengukur jaraknya memberikan persentase hasil DR sebesar 0%, FPR 0,8333%, dan ACC 99,1667%.

Persentase FPR dan ACC menunjukan masih terjadi kesalahan klasifikasi trafik normal flashcrowd sebagai Neptune dan Smurf, karena pada saat-saat tertentu nilai fitur trafik normal flashcrowd mendekati rata-rata profil trafik Neptune atau Smurf. Persentase DR 0% menunjukan bahwa tidak ada pendeteksian serangan karena seperti kita ketahui sebelumnya *flashcrowd* merupakan anomaly traffic normal namun bukan suatu bentuk serangan. Hal tersebut terbukti pada gambar 3.2.2.1 apabila melihat dari jumlah paket per flow antara trafik normal dan flashcrowd keduanya memiliki jumlah yang relatif sama dan tidak terjadi peningkatan jumlah yang tinggi namun pada DDoS terjadi lonjakan nilai, sehingga terlihat grafiknya berbeda dengan normal dan *flashcrowd*.

3.3. Stabilitas Sistem



Gambar 3.3.1 Grafik Jumlah Paket per Flow Trafik

Sistem yang dibangun melakukan proses pada setiap data masukan trafik secara individual hal ini berarti sistem melakukan analisis pada setiap flow trafik masukan. Pada gambar 3.3.1 dapat dilihat bahwa jumlah flow data masukan pada setiap dua detik berbeda, pada trafik normal sendiri pun jumlah flow berubah-ubah, bahkan terjadi kenaikan jumlah flow pada trafik flashcrowd secara gradual dan kenaikan yang tinggi secara tiba-tiba pada trafik DDoS. Saat trafik normal sistem akan memproses kurang lebih sekitar 400 kali proses, saat terjadi anomali sistem melakukan proses 4 kali lebih banyak atau sekitar 2000 kali proses. Hal tersebut berarti proses yang dilakukan oleh sistem tidaklah sama tiap waktunya, waktu yang dibutuhkan sistem untuk menyelesaikan sejumlah proses pun tidak akan sama tiap waktunya. Keadaan seperti itu dapat mengindikasikan sistem kurang stabil.

4. Kesimpulan dan Saran

4.1. Kesimpulan

Kesimpulan dari penelitian tugas akhir ini adalah sebagai berikut:

1. Pendekatan Multivariate Correlation Analysis (MCA) berbasis Triangle-Area-Map (TAM) dapat mendeskripsikan hubungan antar data tiap fitur dalam suatu trafik dalam sistem deteksi anomali
2. Threshold yang terlalu besar dapat menyebabkan kesalahan deteksi dan memperbesar persentase FPR

3. Data masukan homogen atau heterogen tidak mempengaruhi hasil deteksi dan klasifikasi
4. Penghitungan jarak antara TAM data trafik yang diobservasi dengan rata-rata TAM profil trafik acuan dapat digunakan untuk mengklasifikasikan trafik yang diobservasi berdasarkan hasil nilai terkecil perhitungan.
5. Mahalanobis Distance dapat merepresentasikan lebih baik daripada Cosine Distance melihat dari hasil klasifikasi MD memiliki persentase DR lebih tinggi dan FPR lebih rendah daripada CD.
6. Flashcrowd dapat dibuktikan sebagai trafik normal namun merupakan anomali berdasarkan hasil klasifikasi sistem yang menyatakan 99,1667% data flashcrowd yang diuji merupakan trafik normal dan nilai paket/flow antara trafik normal dan flashcrowd yang cenderung sama.

4.2. Saran

Saran untuk penelitian yang selanjutnya antara lain:

1. Melakukan pemilihan fitur yang tepat agar dapat memberikan hasil deteksi dan klasifikasi yang lebih baik.
2. Dapat menemukan metode penghitungan jarak yang lebih baik untuk mengklasifikasikan anomaly traffic.
3. Penelitian selanjutnya diharapkan dapat menggunakan lebih banyak jenis dataset sehingga kemampuan sistem dapat lebih teruji dengan berbagai variasi trafik.
4. Membuat sistem yang lebih stabil dengan kemampuan dan performa yang lebih baik.

Daftar Pustaka

[1] Y. Purwanto, Kuspriyanto, Hendrawan dan B. Rahardjo, "Traffic Anomaly Detection in DDoS Flooding Attack," *International Conference on Telecommunication, System, Services, and Application*, vol. 8, pp. 313-318, 2014.

[2] Z. Tan, A. Jamdagni, X. He, P. Nanda dan R. P. Liu, "A System for Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis," *IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS*, vol. XXV, no. 2, pp. 447-456, February 2014.

[3] J. Sen, "A Robust Mechanism For Defending Distributed Denial of Service Attacks On Web Servers," *International Journal of Network Security & Its Applications (IJNSA)*, vol. III, no. 2, pp. 162-179, March 2011.

[4] P. R. Reddy, R. Siva dan C. Malathi, "Techniques to Differentiate DDOS Attacks from Flash Crowd," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. III, no. 6, pp. 295-299, June 2013.

[5] R. Nagadevi, P. N. Rao dan R. Anand, "A New Way of Identifying DOS Attack Using Multivariate," *International Journal of Computational Engineering Research (IJCER)*, vol. IV, no. 8, pp. 60-64, August 2014.

[6] P. Chitra, P. Pooja, V. Vijayalakshmi dan S. Divya, "Detecting Denial of Service attack using Multivariate Correlation Analysis," *International Journal of Advanced Research in Computer Science Engineering and Information Technology*, vol. II, no. 1, pp. 13-17, February 2014.

[7] "KDD Cup 1999," The UCI KDD Archive Information and Computer Science University of California, Irvine, 28 October 1999. [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html. [Diakses October 2014].

[8] A. O. Adetunmbi, S. O. Adeola dan O. A. Daramola, "Analysis of KDD '99 Intrusion Detection Dataset for Selection of Relevance Features," *World Congress on Engineering and Computer Science*, vol. I, 2010.

[9] M. Arlitt dan T. Jin, "1998 World Cup Web Site Access Logs," August 1998. [Online]. Available: www.acm.org/sigcomm/ITA/. [Diakses October 2014].

[10] W. Wang, S. J. Knapskog dan S. Gombault, "Attribute Normalization in Network Intrusion Detection," *Proc. 10th Int'l Symp. Pervasive Systems, Algorithms, and Networks (ISPAN)*, pp. 448-453, 2009.

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

[12] Z. Tan, A. Jamdagni, X. He, P. Nanda dan R. P. Liu, "Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis," *International Conference on Neural Information Processing (ICONIP)*, vol. XVIII, no. 3, pp. 756-765, 2011.

[13] K.Sujithra dan V. Kumar, "A Survey On Triangle Area Map Based Multivariate Correlation Analysis To Detect Denial-Of-Service Attack," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. III, no. 10, pp. 8203-8206, October 2014.