Analisis Sistem Deteksi Anomali Trafik Menggunakan Algoritma CURE (*Clustering Using Representatives*) dengan Koefisien *Silhouette* dalam Validasi *Clustering*

Analysis of Traffic Anomaly Detection System Using CURE Algorithms (Clustering Using Representatives) with the Silhouette Coefficient for Clustering Validation

Angger Kartyasa Pribadi Putra[1], Yudha Purwanto[2], Astri Novianty[3]

[1,2,3]Prodi S1 Sistem Komputer, Fakultas Teknik Elektro, Telkom University
Bandung, Indonesia
[1]angger@students.telkomuniversity.ac.id , [2]omyudha@telkomuniversity.ac.id , [3]astrinov@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*), *flashcrowd*, dan sebagainya. Namun yang menjadi perhatian dalam analisis proses pengelompokkan anomaly tersebut adalah masalah pelabelan dan validasi tiap objek dari hasil proses *clustering* tersebut.

Dengan memvalidasi *cluster* kita akan mendapatkan jumlah *cluster* optimal dalam analisis anomali trafik dalam hal ini adalah metode *clustering* CURE (*Clustering using Representatives*). Hasil dari validasi akan menjelaskan bagaimana kualitas *cluster* dan tiap objek menggunakan teknik *silhouette index*. Tujuan utama dalam penerapan validasi ini merupakan modifikasi dari algoritma CURE dengan fokus utama yaitu masalah pelabelan tiap objek pada tiap *cluster* dan juga validasi dari hasil *clustering* algoritma CURE.

Hasil dari penelitian ini, algoritma CURE mendapatkan nilai validasi terbaik menggunakan teknik silhouette untuk Dari analisis hasil clustering algoritma CURE didapatkan nilai validasi algoritma CURE menggunakan teknik *silhouette* pada dataset KDDCUP'99 diperoleh nilai rata-rata *silhouette* tertinggi dengan *accuracy* 97.96%, dan nilai rata-rata silhouette *cluster* 0.7642748. Pada dataset Darpa *Week* 5 *Friday* dengan nilai *accuracy* 98.56%, dan nilai rata-rata silhouette *cluster* 0.763525532.

Kata Kunci : anomali trafik, *clustering*, validasi *cluster*, algoritma *CURE, Silhouette Coefficient*

Abstract

In the development of the Internet network technology is now widely discusses phenomena atapun attack threat to a computer or server. Lots of various types of threats on a computer in an Internet network such as DoS (Denial of Service), DDoS (Distributed Denial of Service), a flash crowd, and so on. However, the attention in the analysis process of grouping these anomalies is the problem of labeling and validation of each object on the results of the clustering process.

By validating the cluster we will get the optimal number of clusters in the analysis of traffic anomalies in this case is the method of clustering CURE (Clustering using Representatives). Results of the validation will explain how the quality of the cluster and each object using silhouette technique index. The main objective in the application of this validation is a modification of the algorithm CURE with the main focus of the issue of labeling each object in each cluster and also the validation of the results of CURE clustering algorithm.

Results from this research, algorithm validation CURE get best value using silhouette technique for analysis of the results of clustering algorithms CURE CURE algorithm validation values obtained using the technique dataset KDDCUP'99 silhouette on the average values obtained the highest silhouette with accuracy 97.96%, and the average value 0.7642748 silhouette average cluster. At Darpa dataset Week 5 Friday with a value of 98.56% accuracy, and the average value 0.763525532 silhouette cluster.

Keywords: traffic anomaly, clustering, cluster validation, CURE algorithm, Silhouette Coefficient

1. Pendahuluan

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. Seiring dengan perkembangan teknologi dan informasi, saat ini internet menjadi bagian dari kehidupan masyarakat modern. Internet telah memberikan berbagai kemudahan dengan akses informasi yang luas dan beragam. Dengan perkembangan internet saat ini, akses terhadap internet memicu terjadinya trafik yang tinggi/anomali trafik di dalam suatu jaringan komputer. Trafik yang tinggi/anomali trafik dapat terjadi akibat dua hal, yaitu flash crowd dan serangan flooding trafik.

Flash crowd merupakan fenomena peningkatan akses / trafik secara tinggi ke suatu server karena suatu kejadian tertentu. Kejadian yang menyebabkan terjadinya flash crowd misalnya. Serangan Denial of Service (DoS) dan Distributed Denial of Service (DDoS) merupakan serangan flooding trafik. Serangan Denial of Service (DoS) dan Distributed Denial of Service (DDoS) merupakan serangan yang terjadi dengan membanjiri lalu lintas jaringan dengan banyak data (traffic flooding) atau membanjiri jaringan dengan banyak request terhadap sebuah host atau service sehingga tidak dapat diakses oleh user yang terdaftar / berhak (request flooding). Serangan DoS dan DDoS akan menggunakan resource jaringan yang sangat besar karena paketnya sangat banyak, sehingga sumber daya bandwith pada server penuh kemudian sistem pengolah kehabisan sumber daya dan akhirnya crash / down sehingga tidak dapat melayani servis yang diminta user. Dalam skema serangannya, DDoS akan menyerang sisi availability dari suatu service yang terjadi antara server target serangan dan user yang terdaftar / berhak.

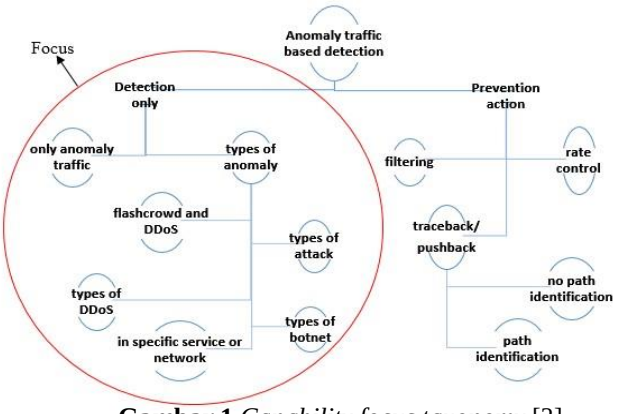
Oleh karena itu, diperlukan suatu sistem untuk mendeteksi anomali trafik di jaringan komputer. Salah satu teknik deteksi yang dapat digunakan untuk mendeteksi anomali trafik adalah clustering based. Clustering memiliki algoritma-algoritma yang diterapkan untuk melakukan proses-prosenya, salah satu algoritma clustering adalah CURE (Clustering Using REpresentatives) yang berbasis hierarchical. Dengan teknik clustering based, maka akan dilakukan penganalisaan data ke dalam struktur kelompok-kelompok data yang memiliki kesamaan berdasarkan jenis serangan (Normal dan Flash crowd) dengan algortima CURE yang digunakan.

2. Dasar Teori

2.1 Deteksi Anomali Trafik

Dalam penelitian dan survey [1] dalam sistem deteksi anomali trafik ini dilakukan suatu pendekatan ke masalah deteksi serangan pada jaringan komputer atau dikenal sebagai Intrusion Detection System (IDS). Intrusion Detection System (IDS) adalah suatu rancangan sistem yang merupakan bagian dari sistem keamanan jaringan komputer untuk menjaga integritas dan melayani ketersediaan layanan bagi seluruh pengguna jaringan komputer. Teknik IDS ini digunakan untuk memantau aktivitas jaringan dalam waktu tertentu dan juga menetapkan suatu nilai (threshold) sebagai parameter acuan (baseline) untuk mendeteksi suatu serangan. Misi dari IDS yaitu mencakup 3 hal penting yaitu kerahasiaan (confidentiality), integritas (integrity), dan ketersediaan (availability). menyatakan bahwa dalam teknik IDS, suatu system tersebut harus memonitor dan mendeteksi adanya serangan (intusi) lalu memberikan tanda (sign) pada sistem administrator bila terjadi serangan untuk ditindak lanjuti.

Pengembangan dari teknik IDS ini menjadi suatu acuan untuk adanya suatu teknik untuk mencegah serangan tersebut atau dikenal sibagai Intrusion Prevent System (IPS) yang memiliki kemampuan dalam memantau dan mendeteksi serangan layaknya IDS dan melakukan aksi dalam mengatasi serangan secara otomatis. Adapun fokus dalam deteksi anomali trafik tercantum dalam taksonomi pada gambar 1. Dibutuhkan beberapa tools dalam mengenali sebuah anomali trafik. Mengenali sebuah anomali bisa dilakukandengan cara analisis secara visualdenganmelihat adanya anomali pada aliran trafik pada jaringan, tentunya dengan bantuan software flow analisis seperti Wireshark. Namun dari data anomali trafik tersebut harus ditentukan fitur-fitur yang akan dianalisis selanjutnya.



Gambar 1 *Capability focus taxonomy* [3]

2.2 Serangan *Flooding Traffic (DDoS)* dan *Flash Crowd*

Dalam sebuah penjelasan [2] sebuah peristiwa flash crowd adalah sebuah lonjakan besar dalam lalu lintas jaringan internet sebagai contoh pada situs Web tertentu menyebabkan peningkatan dramatis dalam beban server dan menempatkan tekanan berat pada *link* jaringan yang mengarah ke server, yang mengakibatkan peningkatan yang sangat derastis dalam packet loss dan tingkat kepadatan suatu trafik. Lonjakan jumlah *user* ini terjadi secara alamiah atas suatu kejadian yang biasanya kenaikannya tidak seketika (*drastic*) tetapi gradual seiring penyebaran informasi kejadian.

Distributed Denial of Service (DDoS) merupakan upaya eksplisit oleh penyerang untuk mencegah pengguna untuk menggunakan suatu jenis layanan [3]. Dalam definisi secara luas yaitu setiap upaya untuk merusak dan menolak segala layanan

yang ada. Contoh serangan dari DdoS yaitu seperti *TCP SYN flooding*, HTTP (*hyper text transfer protocol*) request flooding termasuk serangan untuk melindungi password pada halaman web, atau mencoba membuat server Web menjadi bermasalah (*crash*) seperti baru-baru ini yaitu serangan Code Red

[4]. Menurut survey paper sasaran serangan dalam DoS/DDoS adalah *link*/*bandwidth* untuk membuat sumber daya *bandwidth* penuh atau tidak tersedia lagi bagi user yang akan mengakses. Akibat dari hal ini yaitu dan sumber daya komputasi baik dari proses, *memory*, *buffer* pada *server* maupun *node* jaringan akhirnya menjadi *crash* / *down* sehingga tidak dapat melayani servis yang diminta *user*. Akibatnya koneksi komunikasi tersebut memuat pesan yang besar untuk memenuhi *link* dan juga menghabiskan sumber daya komputasi target sehingga server menjadi *down*/*crash*.

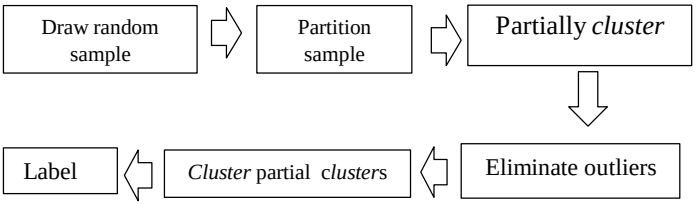
2.3 Algoritma Clustering CURE (Clustering Using Representatives)

Dari penelitian [5], permasalahan umum pada algoritma *hierarchical clustering* tradisional yaitu tidak mendukung untuk menangani dataset dalam ukuran besar. Algoritma CURE (*Clustering Using Representatives*) adalah algoritma *clustering* berbasis *hierarchical clustering* yang efisien untuk *dataset* yang lebih besar. Berikut merupakan prosedur *clustering* algoritma CURE secara umum :

1. Menetapkan jumlah poin perwakilan sasaran c, untuk setiap *cluster*, pilih c titik / poin yang tersebar mencoba untuk menangkap bentuk fisik dan geometri *cluster*.
2. Titik yang tersebar yang dipilih kemudian menyusut menuju *centroid* dalam sebuah nilai (pecahan) kecil dari α dimana $0 < \alpha < 1$.
3. Titik-titik ini digunakan sebagai perwakilan dan pada setiap langkah dari algoritma, dua *cluster* dengan pasangan terdekat dari perwakilan kemudian digabung ($d_{\min}$).
4. Setelah setiap penggabungan, titik / poin c lain dipilih dari perwakilan asli *cluster* sebelumnya untuk mewakili *cluster* baru.
5. Penggabungan *cluster* berhenti sampai target k *cluster* ditemukan.

Berdasarkan penelitian ini, dalam hal menangani *databases* dalam ukuran besar, CURE menerapkan kombinasi dari *random sampling* dan *partitioning*. Pada penelitian ini, CURE menerapkan algoritma hierarchical *cluster*ing dengan mengambil jalan tengah antara the *centroid-based* dan *all-point extremes.* Overview proses

algoritma *clustering* CURE tercantum dalam gambar 2.

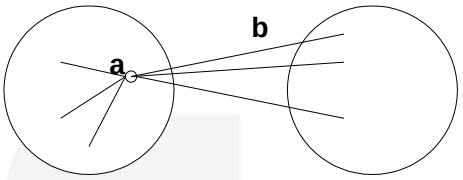


Gambar 2 Overview of Cure Algorithm [5]

2.4 Validasi Klaster

2.4.1 Teknik Silhouette Coefficient

Silhouette Coefficient adalah metode penafsiran untuk validasi klaster pada objek-objek. Teknik ini memberikan representasi grafis singkat tentang seberapa baik setiap objek terletak dalam *cluster*-nya. *Silhouette Coefficient* dikembangkan pertama kali oleh Kaufman dan Rousseeuw [6]. Sebuah nilai koefisien *silhouette* dari sebuah objek semisal Aj berada pada rentang antara -1 sampai dengan 1. Semakin dekat nilai *silhouette* objek Aj ke 1, maka semakin tinggi derajat kepemilikan objek Aj di dalam *cluster* tersebut. Berikut penjelasan dalam menentukan silhouette coefficient dari *cluster* :



Gambar 3 Ilustrasi silhouette coefficient dengan a(i) dan b(i) [6]

Dalam sebuah ilustrasi diberikan pada gambar 3 sebuah titik x di *cluster* A, lalu a(i) adalah jarak rata-rata antara titik x dengan titik lain di dalam *cluster* A, dan b(i) adalah jarak rata-rata antara titik x dan titik-titik di *cluster* kedua yang terdekat dengan A, yaitu *cluster* B. Tahapan perhitungan *Silhouette Coefficient* adalah sebagai berikut :

1. Hitung rata-rata jarak dari suatu objek misalkan i dengan semua objek lain yang berada dalam satu cluster dengan j adalah objek lain dalam satu cluster A dan d(i,j) adalah jarak antar objek i dengan j.

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A,\, j \neq i} d(i,j) \quad (1)\ [6]$$

2. Kemudian hitung juga rata-rata jarak dari objek i tersebut dengan semua objek di cluster lain, lalu diambil nilai terkecilnya.

$$d(i,C) = \frac{1}{|A|} \sum_{j \in C} d(i,j) \quad (2)\ [6]$$

dengan d(i,C) adalah jarak rata-rata objek i dengan semua objek pada cluster lain C dimana $A \neq C$.

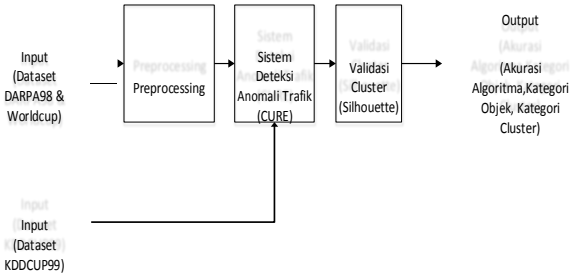
$$b(i) = \min C \neq A\ d(i,C) \quad (3)\ [6]$$

3. Maka Silhouette Coefficient nya adalah:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (4)\ [6]$$

3. Metodologi

3.1 Gambaran Umum Sistem



Gambar 4 Alur proses sistem deteksi anomali trafik

3.2 Dataset

Dataset merupakan data sebagai dasar acuan dan data reset awal yang digunakan pada penelitian ini. Dataset yang digunakan diantaranya KDDCup'98, Darpa'98 dan dataset worldcup'98.

3.3 Akurasi Algoritma CURE

Analisis akurasi dilakukan untuk mengetahui seberapa akurat algoritma modified CURE untuk dapat membedakan antara traffic normal atau serangan. Parameter yang digunakan untuk mengukur hasil deteksi adalah sebagai berikut :

$$\text{Akurasi} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (5)$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (6)$$

$$\text{Precision} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (7)$$

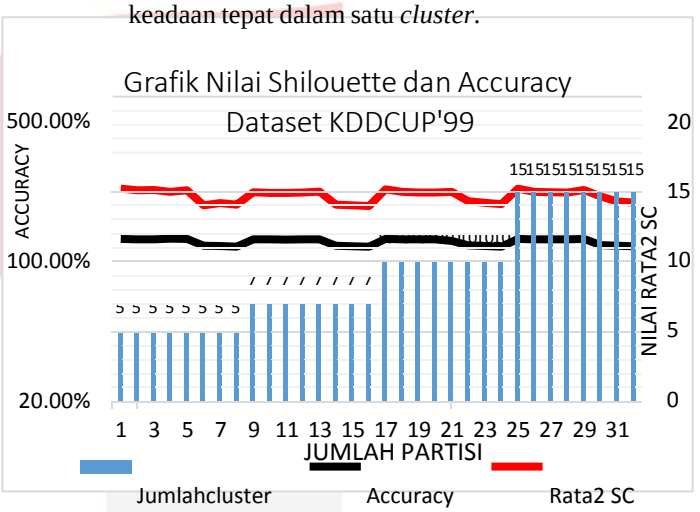
4. Pengujian dan Analisis

4.1 Pengujian dan Analisis Pengaruh Jumlah cluster dan Jumlah Partisi Terhadap Hasil Akurasi dan Validasi cluster Algoritma CURE

Dari hasil percobaan validasi algoritma CURE menggunakan teknik silhouette pada dataset KDDCUP'99 diperoleh nilai rata-rata silhouette tertinggi yaitu dengan jumlah *cluster* 15, dengan jumlah partisi sebesar 10, accuracy 97.96%, dan nilai silhouette cluster 0.7642748. Bila dilihat distribusi serangan terdapat 10 macam jenis serangan. Dari informasi pada gambar 4.1 tersebut diperkirakan rentangan jumlah cluster yang dipakai yaitu lebih dari 5 cluster. Namun dalam algoritma CURE terdapat parameter jumlah partisi yaitu untuk membagi cluster tersebut kedalam beberapa bagian. Pada percobaan diatas diberikan rentangan jumlah partisi antara 10-100 untuk mengetahui apa saja dampak perubahan parameter jumlah cluster dan jumlah partisi terhadap akurasi algoritma CURE dan nilai validasi clustering CURE.

Pada grafik nilai akurasi dan nilai validasi clustering CURE pada gambar 5 terlihat suatu linearitas antara keduanya secara kasat mata. Nilai akurasi algoritma dan nilai silhouette clustering CURE dipengaruhi oleh jumlah cluster yang tepat dan pemilihan jumlah partisi yang tepat. Pada grafik diatas terjadi penurunan

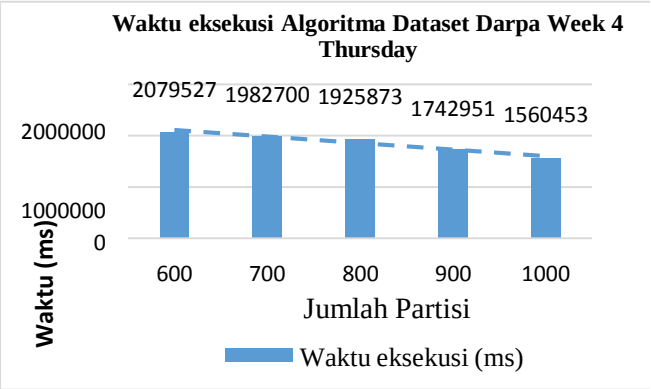
nilai akurasi dan silhouette cluster akibat bertambahnya jumlah partisi yang diberikan. Sebagai contoh pada jumlah cluster 15 mengalami penurunan nilai akurasi algoritma dan nilai silhouette saat jumlah partisi naik yang semula dengan rata-rata 95% menjadi 89%. Kategori cluster yang diperoleh dari hasil percobaan yaitu dari rentangan 0.5-0.7 tersebut menandakan hasil validasi algoritma clustering CURE yaitu sebuah cluster yang kuat atau dengan kata lain nilai objek masih dalam keadaan tepat dalam satu *cluster*.



Gambar 5 Grafik Nilai silhouette dan accuracy dataset KDDCUP'99

4.2 Pengujian dan Analisis Pengaruh Jumlah Partisi Terhadap Waktu Eksekusi Algoritma CURE

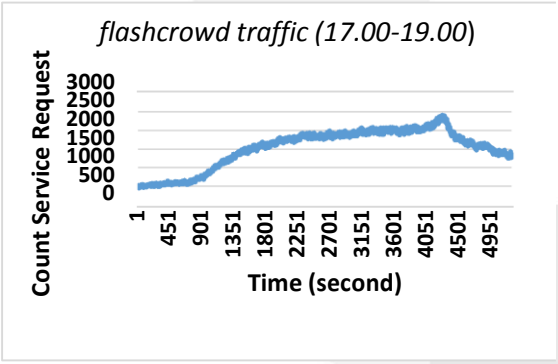
Pada grafik gambar 6 contoh dataset diatas terlihat jelas pengaruh penambahan nilai dari parameter jumlah partisi dari algoritma CURE untuk waktu eksekusi. Pada gambar 6 terlihat bahwa CURE dapat menangani dataset besar (900000 data). Semakin besar nilai partisi yang diberikan maka semakin kecil waktu eksekusi algoritma yang dibutuhkan. Sehingga peningkatan nilai parameter partisi berbanding terbalik dengan waktu eksekusi. Selain itu perbedaan jumlah data juga berpengaruh, semakin banyak jumlah data, semakin banyak pula waktu ekskusi yang diperlukan.



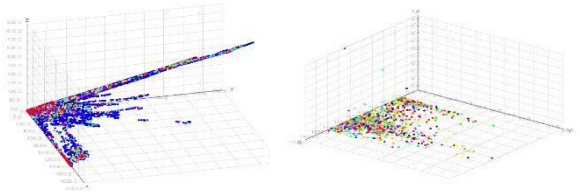
Gambar 6 Grafik Pengaruh Parameter Jumlah Partisi terhadap Waktu Eksekusi Algoritma pada Dataset Normal "Darpa week 4 Thursday"

4.3 Pengujian dan Analisis Data Flashcrowd

Digunakan dataset WorldCup 98 untuk menganalisa bahwa data *flashcrowd* bukan termasuk kedalam data serangan namun termasuk dalam data normal namun dalam request yang tinggi dalam jumlah request service. Total data pada WC 98 sebanyak 5.400 data normal, 5400 data saat flashcrowd dan 10.800 data gabungan normal dengan flashcrowd. Dataset worldcup 98 yang dipakai pada tugas akhir ini memilih data yang memiliki tingkat kenaikan trafik/lonjakan *request* yang paling signifikan yaitu pada hari ke 66 pada jam 17.00-19.00 yang ditunjukkan
pada Gambar 7 itulah yang disebut dengan *flashcrowd*.



Gambar 7 Laju Trafik WorldCup 98 hari 66 (17.00-19.00)



Gambar 8 Hasil plotting *cluster* dataset serangan "KDDCUP'99" (kiri) dan dataset *flash-crowd* (kanan)

Pada Gambar 7 menunjukkan jumlah *flow* dan *request* terjadi peningkatan secara *gradual* (perlahan) pada saat *flashcrowd*. Namun apabila ingin melihat apakah flashcrowd adalah serangan atau bukan, hasil plotting *cluster* serangan dan flashcrowd pada Gambar 8 menunjukkan bahwa plotting flashcrowd berkerumun dibawah dengan jumlah yang banyak dibandingkan dengan dataset yang mengandung serangan terdapat kenaikan di dalam hasil plotting tersebut. Hal ini dapat dikatakan bahwa *flashcrowd* memang sebuah trafik normal yang anomali namun di dalamnya terdapat banyak sumber permintaan (*request*) secara gradual dalam sebuah waktu/kejadian.

Tabel 1 Hasil pengujian *flashcrowd*

		Aktual	
		Positive	Negative
Prediksi	Positive	TP = 0	FP = 0
	Negative	FN = 0	TN = 216

Accuracy = 100%

Pada tabel 1 terbukti data *flashcrowd* pada hari ke 66 merupakan data normal yang terdeteksi dengan nilai TN yang di prediksi adalah data normal mempunyai nilai aktual berupa data normal juga, namun terjadi kenaikan trafik yang signifikan. Karena seperti yang kita ketahui sebelumnya pada asumsi di atas bahwa *flashcrowd* merupakan anomali trafik normal bukan suatu bentuk serangan. Pengujian ini dilakukan untuk membuktikan asumsi bahwa *flashcrowd* bukanlah sebuah serangan.

5. Kesimpulan

5.1 Kesimpulan

1. Hasil akurasi algoritma CURE dalam mengatasi anomali trafik dalam dataset KDDCUP'99 yaitu mencapai nilai 97.96%, Detection Rate 100% ,dan False Positive Rate 0% disimpulkan bahwa sistem dapat mendeteksi anomali trafik dengan baik .

2. Dari analisis hasil clustering algoritma CURE didapatkan nilai validasi algoritma CURE menggunakan teknik silhouette pada dataset KDDCUP'99 diperoleh nilai rata-rata silhouette tertinggi dengan jumlah cluster 15, jumlah partisi 10, accuracy 97.96%, dan nilai rata-rata silhouette cluster 0.7642748 disimpulkan bahwa sistem dapat mendeteksi anomali dengan baik dengan kategori cluster yang baik.

3. Pada dataset Darpa Week 4 Thursday, pada saat diberikan nilai partisi dari 600 dan 100 partisi dengan waktu eksekusi algoritma yang diberikan yaitu 2079527 ms dan 1560453 ms dapat disimpulkan bahwa semakin besar nilai partisi yang diberikan maka semakin kecil waktu eksekusi algoritma yang dibutuhkan.

4. Pada dataset worldcup'98 pada hari ke 66 terdeteksi dengan nilai True Negatif yang di prediksi adalah data normal mempunyai nilai aktual berupa data normal juga, namun terjadi kenaikan trafik yang signifikan dapat disimpulkan bahwa flashcrowd merupakan anomali trafik normal bukan suatu bentuk serangan.

5.2 Saran

1. Proses *clustering* Algoritma CURE, dipengaruhi dengan penggunaan jarak menggunakan *Euclidian Distance*, untuk meningkatkan tingkat disamiliritas yang lebih baik dari sebelumnya dapat dikembangkan dengan rumus selain *Euclidian Distance*.
2. Untuk teknik silhouette coefficient yang dipengaruhi model persebaran data, dapat diteliti jauh lebih lagi ke dalam dataset yang berbeda dengan melihat keberagaman model persebaran data yang ada dalam kondisi real.
3. Keterbatasan validasi silhouette coefficient yaitu terdapat pada peninjauan semua jumlah data hasil *clustering*, dikarenakan pertimbangan semua titik objek yang telah *tercluster*.

Referensi

[1] V. L. L. Thing, M. Sloman and N. Dulay, "A Survey of Bots Used for Distributed," May 2007.

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

[3] T. Velmurugan and T. Shantanam, "A Survey of Partition Based of Clustering Algoritms in Data Mining : An Experimental Approach," *Information Technology Journal*, pp. 478-484, 2011.

[4] Gerard Munz, Sa Li, Georg Carle, "Traffic Anomaly Detection Using K-Means Clustering," in *GI/ITG-Workshop MMBnet*, Hamburg, Germany, 2007.

[5] 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.

[6] "Denial of Service Attacks," CERT Coordination Center, 1999. [Online]. Available: http://www.cert.org/tech tips/denial of service.html. [Accessed 25 September 2014].

[7] "TCP SYN Flooding and IP Spoofing," CERT Advisory, September 1996. [Online]. Available: http://www.cert.org/advisories/CA-1996-21.html. [Accessed 25 September 2014].

[8] D. Moore, "The Spread of the Code-Red Worm (CRv2)," August 2001. [Online]. Available: http://www.caida.org/analysis/security/code-red/coderedv2 analysis.xml. [Accessed 25 September 2014].

[9] Martin Arlitt and Tai Jin, Workload Characterization of The 1998 World Cup Web Site, HP Laboratories Palo Alto, 1999.

[10] Rui Xu and Donald Wunsch II, "Survey of Clustering Algorithms," *IEEE TRANSACTIONS ON NEURAL NETWORKS*, vol. 16, no. 3, pp. 645-678, MAY 2005.

[11] L.kaufmann, "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis," *Computational and Applied Mathematics*, 1987.