ANALISIS SISTEM DETEKSI ANOMALI TRAFIK MENGGUNAKAN ALGORITMA *CLUSTERING* CURE (*CLUSTERING USING REPRESENTATIVES*) DENGAN *OUTLIER REMOVAL CLUSTERING* DALAM MENANGANI *OUTLIER*

ANALYSIS OF TRAFFIC ANOMALY DETECTION SYSTEM USING CURE (CLUSTERING USING REPRESENTATIVES) CLUSTERING ALGORITHM WITH OUTLIER REMOVAL CLUSTERING FOR HANDLING OUTLIERS

Muhammad Agung Tri Laksono[1], Yudha Purwanto[2], Astri Novianty[3]

Telkom University
Bandung, Indonesia aciilll@students.telkomuniversity.ac.id[1],
omyudha@telkomuniversity.ac.id[2], astrinov@telkomuniversity.ac.id[3]

Abstrak

Perkembangan pesat teknologi dan informasi khusunya internet sekarang ini memicu munculnya fenomena-fenomena anomali trafik (serangan) atapun ancaman terhadap sebuah komputer atau server. *Flash crowd* merupakan fenomena peningkatan akses / trafik secara tinggi ke suatu *server* karena suatu kejadian tertentu. 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* (*request flooding*) sehingga tidak dapat diakses oleh user yang terdaftar / berhak *(legitimate user)*. Oleh karena itu, perlu adanya suatu sistem deteksi dengan melakukan pengelompokan pada anomali trafik.

Pada penelitian Tugas Akhir ini digunakan salah satu teknik dalam deteksi anomali trafik yaitu *clustering based*. Algoritma CURE merupakan salah satu algoritma *clustering* berbasis *hierarchical* yang memiliki prestasi dapat mengatasi *outlier* Kemudian, fokus penelitian Tugas Akhir ini adalah dalam hal menangani titik *outlier* dari dataset yang digunakan. *Outlier* dieliminasi dengan menghapus titik yang dianggap sebagai *outlier* dengan teknik *outlier removal clustering (ORC)*.

Hasil dari penelitian ini, algoritma CURE memiliki perfomansi yang baik dalam mendeteksi anomali trafik. Hal itu dapat ditunjukkan dengan pengujian yang dilakukan dengan dataset DARPA 1998, dimana nilai rata-rata *detection rate* sebesar 98.4588 %, *false positive rate* 0.2325 %, dan *accuracy* 94.7323 %. Hasil pengujian eliminasi *outlier* dengan *threshold* 0.1 – 1, teknik ORC berhasil menemukan dan menghapus titik yang dianggap sebagai *outlier*.

Kata kunci: anomali trafik, ddos, *flash crowd, preprocessing, clustering,* algoritma cure

Abstract

The rapid development of information technology and especially the Internet today triggers phenomena traffic anomalies (attacks) or threats to a computer or server. Flash crowd is a phenomenon of increasing access / traffic is high to a server for a particular event. Denial of Service (DoS) and Distributed Denial of Service (DDoS) is an attack that occurred with flood the network with a lot of data (traffic flooding) or flood the network with a lot of requests to a host or service that can't be accessed by registered users (request flooding). Therefore, we need a system of detection by clustering the traffic anomalies.

This final project research used a technique to detect the traffic anomaly which is clustering based. CURE algorithm is a hierarchical based clustering algorithm which has ability in terms of handling outliers. Then, the focus of this final project research is in terms of handling the outliers points from the dataset. Outliers eliminated by removing a point that was considered as an outlier with outlier removal technique clustering (ORC).

The results from this study, CURE algorithm has a good performance in detecting anomalous traffic. It show with the tests performed by DARPA 1998 dataset, where the average value of 98.4588 % detection rate, 0.2325 % false positive rate, and 94.7323 % accuracy. The test results of elimination of outliers with threshold value 0.1 - 1, ORC technique successfully found and remove the points that are considered as an outlier.

Keyword: traffic anomaly, ddos, flash crowd, preprocessing, clustering, cure algorithm

1. Pendahuluan

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

Flash crowd merupakan fenomena peningkatan akses / trafik secara tinggi ke suatu server karena suatu kejadian tertentu. 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 (request flooding)* terhadap sebuah *host* atau *service* sehingga tidak dapat diakses oleh *user* yang terdaftar/berhak *(legitimate user)*.

Oleh karena itu, diperlukan suatu sistem untuk mendeteksi *(Intrussion Detection System)* anomali trafik di jaringan komputer. Salah satu teknik deteksi yang dapat digunakan untuk mendeteksi anomali trafik yaitu *clustering based* yang merupakan metode dalam data mining . Pada Tugas Akhir ini, algoritma *clustering* CURE *(Clustering Using Representatives)* yang merupakan modernisasi dari algoritma hirarki tradisional digunakan untuk melakukan proses deteksi. Kemudian, fokus penelitian Tugas Akhir ini adalah dalam hal menangani titik *outlier* dari dataset yang digunakan. *Outlier* dieliminasi dengan teknik *outlier removal clustering (ORC)*. Dengan teknik *clustering based*, maka dilakukan proses penganalisaan data ke dalam struktur kelompok-kelompok data yang memiliki kesamaan berdasarkan jenis anomali trafik (normal, *flash crowd*, serangan *DDoS*) dengan algortima CURE yang digunakan.

2. Tinjauan Pustaka

2.1 Deteksi Anomali Trafik

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). 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. Pada penelitian ini melakukan proses *Intrussion Detection System (IDS)* dalam mendeteksi anomali trafik berupa serangan DDoS dan *flash crowd*. Proses deteksi digunakan menggunakan dataset DARPA 1998 dapat digunakan untuk evaluasi *Intrussion Detection System* (IDS) [2]

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 drastis 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].

2.3 Algoritma Clustering CURE (Clustering Using Representatives)

Dalam penelitian [5] algoritma CURE (*Clustering Using Representatives*) adalah algoritma clustering berbasis *hierarchical clustering* yang efisien untuk *database* besar (dataset) yang lebih kuat terhadap *outliers* dan mengidentifikasi *cluster* yang memiliki bentuk *non-spherical* dan variasi ukuran data yang besar. Algoritma CURE merupakan modernisasi dari algoritma *hierarchical clustering* tradisional yang menerapkan teknik *agglomerative (bottom-up)*, yaitu menggabungkan beberapa *cluster* hingga menjadi satu.

Berikut merupakan prosedur *clustering* algoritma CURE secara umum :

1. Menetapkan jumlah poin perwakilan sasaran c, untuk setiap cluster, pilih c titik / poin yang tersebar untuk menentukan bentuk fisik dan geometri cluster.
2. Titik yang tersebar yang dipilih kemudian dilakuan proses penyusutan 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.

2.4 Teknik Outlier Removal Clustering (ORC)

Outlier adalah penyimpangan data yang terlalu jauh dari data yang lainnya atau dapat diartikan sebagai data pencilan dalam suatu rangkaian data . Adanya data *outlier* ini akan membuat analisis terhadap serangkaian data menjadi bias, atau tidak mencerminkan fenomena yang sebenarnya. Istilah *outlier* juga sering dikaitkan dengan nilai esktrim, baik ekstrem besar maupun ekstrim kecil yang dapat mengganggu distribusi data.

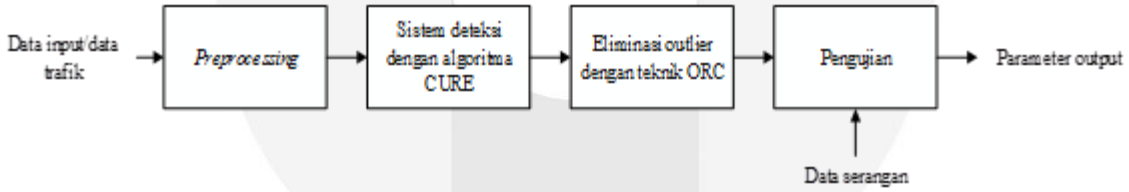
Dalam pengimplementasian *Outlier Removal Clustering* (ORC) terdiri dari 2 tahap. Pertama adalah implementasi algoritma *clustering* yang digunakan, kemudian tahap kedua yaitu secara iterative mengimplementasikan teknik ORC untuk menemukan objek *outlier*. Dalam hal ini *outlier* didefinisikan sebagai titik atau objek yang jaraknya relative berjauhan dari objek-objek lainnya. Teknik ORC ini bergantung pada 2 parameter, yaitu *threshold* dan *distortion*. Threshold merupakan parameter inputan dari pengguna, yang berkisar antara nilai 0 sampai 1, sedangkan *distortion* merupakan perbandingan antara jarak objek terdekat dari *centroid cluster* dan jarak terjauh dari *centroid cluste*r cluster dalam satu cluster [6]. *Distortion* dihitung untuk mengetahui seberapa jauh suatu objek yang terjauh jika dibandingkan dengan objek yang terdekat dengan *centroid*.

Sebuah objek dikatakan *outlier* jika mendapat kondisi *distortion* lebih rendah daripada threshold yang diinputkan, maka objek terjauh dari *cluster* tersebut dapat dipertimbangkan sebagai objek *outlier*. Namun dampak dari teknik ini adalah belum diketahui secara pasti nilai *threshold* yang mengindikasikan *outlier*, karena tiap bentuk *cluster* yang dihasilkan memungkinkan memiliki perbedaan nilai *threshold*.

3. Perancangan Sistem

3.1 Gambaran Umum Sistem

Proses sistem deteksi anomali trafik dilakukan dalam tahap penelitian seperti pada gambar 3.1



Gambar 3.1 Alur proses sistem deteksi

3.2 Data Trafik (Dataset)

3.2.1 DARPA 1998 Dataset

DARPA 1998 merupakan dataset yang dibuat oleh grup *Cyber Security and Information Systems Technology Group* dari MIT Lincoln Laboratory, dibawah *Defense Advanced Research Projects Agency (DARPA ITO)* dan *Air Force Research Laboratory (AFRL / SNHS)* untuk melakukan evaluasi *Intrussion Detection System* jaringan komputer [7]. Dataset ini berisi rekaman trafik normal dan serangan DDoS dalam bentuk *network log connection*. Terdapat dua bagian dalam DARPA 1998 evaluasi intrusi deteksi, yaitu evaluasi *off-line* dan evaluasi *real-time* [7]. Pada penelitian

Tugas Akhir ini, digunakan training data dalam evaluasi *real-time* untuk diolah pada proses *preprocessing* sebelum melakukan proses deteksi anomali trafik menggunakan algoritma CURE.

3.3 Preprocessing

Preprocessing adalah suatu strategi yang saling berkaitan untuk membuat data lebih mudah/cocok untuk digunakan pada data mining. *Preprocessing* merupakan proses untuk mendapatkan suatu data baru dengan melakukan *preprocessing* dari suatu dataset, dimana hasil dari *preprocessing* kemudian digunakan sebagai data input pada proses deteksi anomali trafik.

Pada penelitian Tugas Akhir ini, dibentuk 9 fitur dari dataset DARPA 1998 pada proses *preprocessing* yang dilakukan. Berikut merupakan 9 fitur yang dibentuk dalam periode waktu 1 detik :

Tabel 3.1 Fitur trafik DARPA 1998

Nama Fitur	Kondisi	Keterangan
Count		Jumlah request per detik
IP_Source	Fitur 2-7 berikut ini berdasarkan koneksi IP Source ke IP Dest yang sama	Jumlah koneksi
Protocol		Jumlah protokol yang sama
SYN		Jumlah adanya SYN
ACK		Jumlah adanya ACK
Port_Out		Jumlah port tujuan yang sama
Length		Jumlah length yang sama
Different	Berdasarkan IP Dest yang sama	Kemunculan IP Source yang berbeda tetapi IP Dest sama
New_IP		Kemunculan IP baru (dicek dari IP history, jumlah max history 1000 IP, jika tidak ada dalam list IP history dianggap IP baru)

3.4Sistem Deteksi dengan Metode Clustering

3.4.1 Algoritma Clustering CURE dengan Teknik ORC

Dengan random sample, sebagian besar *outlier* tersaring, selain itu CURE memiliki skema eliminasi *outlier* dalam dua fasa [5]. Fasa pertama berada pada permulaan proses dan fasa kedua berada diakhir proses algoritma. Proses eliminasi pada CURE bergantung dengan nilai *fraction (reducing factor)* pada tiap partisi dan *threshold* jumlah anggota dalam satu *cluster*. Nilai *reducing factor* menentukan gabungan *cluster* yang terbentuk pada tiap partisi. Kemudian, mekanisme eliminasi *outlier* pada CURE yaitu dengan berdasarkan jumlah anggota yang ada dalam *cluster* tersebut. Misal, ditentukan threshold bernilai 1 atau 2, maka setelah proses penggabungan di tiap partisi *cluster* yang memiliki anggota 1 sampai 2 dihapus sehingga dihasilkan *cluster* yang berisi banyak anggota atau setidaknya terdapat 3 anggota dalam *cluster*.

Penelitian Tugas Akhir ini menerapkan metode dari *outlier removal clustering* dalam menangani *outlier* yang terdapat pada dataset. Skema teknik ORC ini juga akan diterapkan pada dua fasa dalam algoritma CURE. Berbeda dengan skema eliminasi *outlier* pada CURE , yaitu dengan berdasarkan *threshold* jumlah anggota dalam *cluster*, pada penelitian ini skema eliminasi *outlier* dilakukan berdasarkan jarak maksimum titik/data dengan titik perwakilan yang ada dalam *cluster*.

Keterangan : untuk fasa ke-1 (9 - selesai kecuali langkah 10) dan ke-2 (10 - selesai) memiliki langkah yang sama hingga akhir proses

Algoritma 1. ORC dalam CURE (fasa ke-1 dan fasa ke-2)

```
1.  Input nilai threshold (Th)
2.  Titik perwakilan = P1
3.  Titik dalam cluster = P2
4.  Titik outlier = P3
5.  Total titik outlier sementara = tmp
6.  Partisi = p;
7.  jarakMin = 1000000
8.  jarakMax, jarakPalingMax = 0;
9.      for semua cluster ∈ semua(p) do (fasa ke-1)
10.       for semua finalMergedCluster do (fasa ke-2)
11.         for 1 to total(P1) do
12.           for 1 to total(P2) do
13.           jarak = euclidean distance(P1,P2)
14.             if jarakMin > jarak then
15.                 jarakMin = jarak
16.             end if
17.             if jarakMax < jarak then
18.                 jarakMax = jarak;
19.             end if
20.                 distorsi = jarakMin / jarakMax
21.           end for
22.         end for
23.           if distorsi < threshold then
24.             if jarakMax > jarakPalingMax then
25.               jarakPalingMax = jarakMax;
26.             end if
27.                   for 1 to total(P1) do
28.                     for 1 to total(P2) do
29.                         jarak = euclidean distance(P1,P2)
30.                         if jarak == jarakPalingMax then
31.                             outlier = P3
32.                             tmp.add(P3)
33.                     end for
34.                   end for
35.           for 1 to tmp do
36.             removeAll(tmp)
37.           end for
38.         end for
39.       end for
40.     end for
```

4. Pengujian dan Analisis

4.3.1 Pengujian dan Analisis Clustering untuk Deteksi Anomali Trafik

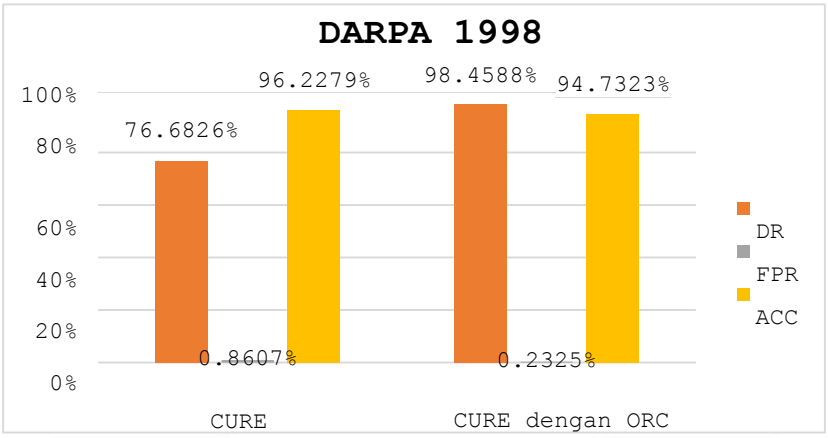
a. Normal + Serangan (DARPA 1998)

Tabel 4.1 Hasil pengujian *Week 1 Tuesday*

Aktual	Prediksi	
	Serangan	**Normal**
Serangan	Total TP = 2	Total FN = 1054
Normal	Total FP = 0	Total TN = 14057

Total TP	Total FP	Total TN	Total FN	DR	FPR	ACC
0.0132 %	0 %	93.0126 %	6.9741 %	100 %	0 %	93.0259 %

Dari hasil pengujian, terlihat bahwa dengan dataset yang digunakan CURE memiliki perfomansi yang baik dalam melakukan deteksi anomali trafik. Hal itu dapat dilihat dari hasil TP dan FP, setelah dilakukan analisa dengan total jumlah data pada tabel dan persentase tersebut algoritma berhasil mendeteksi satu *cluster* yang berisi serangan dan *cluster* tidak memiliki kesalahan deteksi trafik normal sebagai serangan sehingga memiliki nilai DR sebesar 100 % dan FPR sebesar 0 %. Dari hasil TN dan FN, CURE mampu mendeteksi mayoritas *cluster* yang berisi trafik normal dengan jumlah yang lebih banyak dari pada trafik serangan.

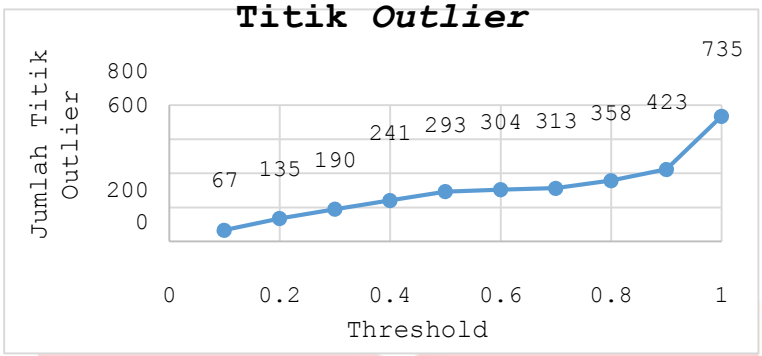


Gambar 4.1 Perbandingan Performansi DARPA 1998

Grafik diatas menggambarkan perbandingan performansi antara CURE dengan metode yang terdapat di dalamnya untuk menangani *outlier* dan CURE dengan teknik ORC untuk proses eliminasi *outlier*. Dengan dataset DARPA 1998, hasil rata-rata DR menunjukkan bahwa CURE dengan ORC memiliki kemampuan deteksi anomali trafik yang lebih baik. Penghapusan titik berdasarkan jarak maksimum yang dilakukan oleh ORC membuat CURE mampu mendeteksi anomali trafik (serangan) dengan lebih baik, berbeda dengan metode eliminasi *outlier* pada CURE dengan melakukan penghapusan *cluster* (sesuai dengan threshold anggota *cluster*). Hal ini menyebabkan kemungkinan *cluster* yang berisi trafik serangan terhapus sehingga berpengaruh terhadap hasil DR. Nilai rata-rata FPR yang dihasilkan menunjukkan baik CURE tanpa ORC maupun CURE dengan ORC memiliki tingkat kesalahan deteksi yang cukup rendah, ini berarti kedua mekanisme eliminasi *outlier* yang berbeda ini memiliki kesalahan deteksi trafik normal sebagai serangan.

Dari hasil *accuracy* dengan dataset DARPA 1998, terlihat pada grafik diatas bahwa CURE tanpa ORC memiliki cenderung memiliki *accuracy* yang lebih baik dari CURE dengan ORC. Hal ini dikarenakan banyaknya jumlah data yang dihapus dengan metode eliminasi *outlier* CURE menghasilkan *cluster* dengan rendahnya kemungkinan tingkat kesalahan deteksi trafik normal sebagai serangan atau sebaliknya. Dengan hal itu maka CURE cenderung memiliki *accuracy* yang baik. Akan tetapi, dari sisi jumlah data yang hilang/*packet loss* pada CURE tanpa ORC sangat tinggi. Hal ini sewaktu-waktu dapat menyebabkan sistem tidak dapat mendeteksi anomali trafik dengan baik atau dengan kata lain memiliki nilai DR yang cukup rendah.

4.3.2 Pengujian dan Analisis Proses Eliminasi Outlier dengan Teknik ORC



Gambar 4.2 Jumlah Titik *Outlier*

Dari hasil pengujian diatas, dari grafik terlihat dengan nilai threshold yang digunakan teknik ORC mampu menemukan *outlier* dalam jumlah yang cukup banyak. Dari total 21.713 random sampel data yang diolah, rata-rata terdapat 305.9 titik yang dianggap sebagai *outlier* dengan nilai *threshold* antara 0.1 – 1. Setelah dilakukan analisa, data trafik normal lebih banyak yang dianggap sebagai *outlier* dari pada trafik serangan. Hal itu karena pada setiap iterasi ORC suatu objek yang memiliki jarak terjauh dari *centroid cluster*, maka objek tersebut dianggap sebagai *outlier*, sehingga tidak dapat dipastikan bahwa data *outlier* merupakan data trafik serangan tetapi ditentukan oleh objek yang memiliki jarak terjauh dari *centroid cluster*.

Terlihat bahwa semakin besar nilai *threshold* maka jumlah titik yang dianggap sebagai *outlier* juga semakin besar. Setelah dilakukan analisa, dengan nilai *threshold* yang semakin besar maka semakin banyak juga nilai distorsi dengan kondisi yang lebih kecil dari *threshold*. Pada perhitungan distorsi di setiap iterasi yang dilakukan, jika jarak antara titik dengan titik perwakilan merupakan jarak yang terjauh, maka titik tersebut dianggap sebagai titik *outlier* dan nilai distorsi semakin kecil. Dengan hal itu maka jumlah distorsi yang memiliki nilai lebih kecil dari *threshold* yang digunakan menyebabkan semakin banyaknya titik yang dianggap sebagai *outlier*. Pada dataset ini, data dianggap sebagai *outlier* dengan rata-rata jarak maksimum sekitar 80 – 200 dari setiap *threshold* yang digunakan pada pengujian. Penghapusan *outlier* dengan teknik ORC yang digunakan tidak berpengaruh terhadap hasil performansi algoritma CURE seperti yang terlihat pada tabel diatas. CURE dengan ORC tetap memiliki performansi yang cukup stabil dalam mendeteksi anomali trafik.

Pada CURE dengan ORC jumlah *outlier* yang dihapus tidak terlalu besar. Perbedaan mekanisme dalam pencarian titik outlier menentukan banyaknya jumlah titik yang dianggap sebagai *outlier* untuk kemudian dilakukan penghapusan. Berdasarkan pencarian jarak maksimum dengan *threshold* yang diinputkan *user*, skema ini membuat CURE dengan ORC dapat melakukan proses pencarian titik *outlier* tanpa mempengaruhi kemampuan deteksi sistem.

5. Kesimpulan

Kesimpulan yang dapat diambil dari penelitian Tugas Akhir ini yaitu :

1. Algoritma CURE secara garis besar memiliki performansi yang relatif baik dalam melakukan deteksi anomali trafik pada semua skema pengujian dengan semua dataset yang digunakan.
2. Performansi sistem deteksi anomali trafik menggunakan dataset DARPA 1998 menunujukan hasil yang yan baik, hal ini menununjukan fitur-fitur yang dihasilkan pada *preprocessing* dataset DARPA 1998 dapat menggambarkan fenomena kejadian trafik anomali dan normal yang kemudian dapat dilakukan proses deteksi dengan baik oleh CURE.
3. Teknik ORC yang diterapkan dalam CURE dapat menemukan titik yang dianggap sebagai *outlier* berdasarkan acuan nilai threshold yang digunakan (0.1 – 1).
4. Semakin besar nilai threshold maka semakin banyak titik outlier yang ditemukan, karena jumlah distorsi dengan kondisi lebih kecil dari threshold semakin banyak. Nilai distrosi bergantung pada perhitungan jarak antara titik dengan titik perwakilan dalam satu *cluster*.
5. Tidak dapat dipastikan bahwa data outlier merupakan data trafik normal atau serangan tetapi ditentukan oleh objek yang memiliki jarak terjauh dari *centroid cluster*.

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] "1998 DARPA Intrussion Detection Evaluation Data Set," LINCOLN LABORATORY MASSACHUSSETS INSTITUTE OF TECHNOLOGY, [Online]. Available: http://www.ll.mit.edu/ideval/data/1998data.html. [Accessed 26 5 2015].

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

[4] "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].

[5] 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].

[6] Sudipto Guha, Rajeev Rastogi, and Kyuseok Shim, "CURE: An Efficient Clustering Algorithm for Large Databases," *In: Proceedings of the ACM SIGMOD International Conference on Management of Data ,* pp. 73-84, 1998.

[7] Ville Hautamaki, Svetlana Cherednichencko, Ismo Karkkainen, Tomi Kinnunen, and Pasi Franti, "Improving K-Means by Outlier Removal," pp. 978-987, 2005.