PERANCANGAN DAN ANALISIS DETEKSI ANOMALI BERBASIS *CLUSTERING* MENGGUNAKAN ALGORITMA *MODIFIED K-MEANS* DENGAN *TIMESTAMP INITIALIZATION* PADA *SLIDING WINDOW*

DESIGN AND ANALYSIS OF ANOMALY DETECTION BASED CLUSTERING USING MODIFIED K-MEANS ALGORITHM WITH TIMESTAMP INITIALIZATION ON SLIDING WINDOW

I Wayan Oka Krismawan Putra[1], Yudha Purwanto[2], Fiky Yosef Suratman[3]

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

Abstrak

***Anomaly Traffic* yang terjadi di *Internet* biasanya membuat pengguna resmi tidak dapat mengakses dengan baik. Hal ini dapat disebabkan oleh peningkatan jumlah pengguna pada satu waktu atau karena serangan *botnet* ke jaringan. Tujuan penelitian ini metode untuk mendeteksi apakah ada *anomaly traffic* atau tidak. Penelitian ini menggunakan algoritma k-means sebagai algoritma deteksi yang dimodifikasi pada penentuan centroid dan inisialisasi *cluster*, di mana inisialisasi *cluster* digunakan *Timestamp* Inisialisasi yang diterapkan dalam penentuan centroid dan cluster berdasarkan titik data point yang diperoses. Modifikasi k-means menggunakan *Timestamp* Inisialisasi dapat menghilangkan penentuan k-value *cluster* yang mempengaruhi detection rate dan false positive rate saat menggunakan berbeda k-value cluster. Penelitian ini juga menggunakan teknik windowing untuk mendapatkan proses yang lebih baik efisien untuk mendeteksi anomaly traffic dengan hasil *detection rate* yang tertinggi 96.06% dan *false positive rate* terkecil 0.75% dari pengujian beberapa *dataset*.**

Kata kunci: Anomaly Traffic, Clustering, Modified K-Means, Timestamp Initialization, Windowing

Abstract

Traffic anomalies that occur on the Internet usually make authorized users cannot access properly. This can be caused by an increase in the number of users at a time or due to the attack of botnet to the network. This research purpose a method to detect whether there is anomaly or not. This research used K-Means algorithm as the detection algorithm that modified on determination of the centroid and the cluster initialization, where the cluster initialization was used Timestamp Initialization as applied which in the determination of the centroid and the cluster based on the processed data point. Modified K-Means using Timestamp Initialization can eliminate the determination of K-value cluster that affect detection rate and false positive rate when using different K-value cluster. This research also used windowing technique to obtain a better efficient processes to detect anomaly traffic with best detection rate is 96.06% and false positive rate is 0.75% from testing with any dataset.

Keyword: Anomaly Traffic, Clustering, Modified K-Means, Timestamp Initialization, Windowing

1. Pendahuluan

Anomaly traffic yang terjadi pada jaringan internet biasanya membuat pengguna resmi tidak dapat melakukan akses sebagaimana mestinya. Hal tersebut bisa dikarenakan oleh peningkatan jumlah pengguna pada suatu waktu atau karena adanya serangan dari pihak lain terhadap jaringan tersebut. Serangan ini bisa berupa serangan DOS maupun DDoS. Contohnya serangan DDoS adalah melakukan pembanjiran jaringan target dengan mengirimkan pesan "PING" sehingga target tidak dapat mengakses ke jaringan secara resmi [1]. Ada dua metode untuk mela kukan deteksi anomaly yaitu dengan signature dan anomaly based. Metode Signature menggunakan database dalam pengenalan suatu anomaly, sehingga jika terdapat anomaly baru tidak akan terdeteksi. Sedangkan, Anomaly tidak menggunakan database tetapi menggunakan pembelajaran pola yang terjadi, sehingga jika terdapat anomaly baru akan bisa terdeteksi.

Anomaly based memiliki kekurangan yaitu di False Detection yang rendah, maka kami menggunakan metode Clustering yang merupakan salah satu metode dari data mining [2] [3], dimana teknik ini melakukan pengelompokan data berdasarkan similaritas suatu data dengan data lainnya. Data dengan similaritas tinggi akan dikelompokan kedalam satu kelompok, dan similaritas rendah akan dikelompokan kedalam kelompok lain. Clustering digunakan untuk mendapatkan informasi baru yang nantinya digunakan sebagai bahan pertimbangan untuk mencapai hasil yang diinginkan

Salah satu algoritma clustering adalah k-means. Beberapa penelitian sebelumnya juga sudah menerapkan k-means sebagai algoritma untuk menentukan sistem deteksi. Prinsip k-means adalah mengelompokkan data berdasarkan similaritas data dengan data lainnya untuk membentuk cluster, dengan similaritas yang tinggi pada sebuah cluster dan similaritas rendah pada cluster yang berbeda. Dalam proses, pertama kita harus menentukan centroid dan jumlah cluster yang akan digunakan untuk menentukan data. Dengan demikian, detection rate dan false positive rate yang dihasilkan tergantung pada centroid dan jumlah cluster awal yang ditentukan. Dengan demikian, dalam jurnal ini kami akan mempurpose:

- Detection anomaly traffic menggunakan TimeStamp Initialization yang diterapkan pada k-means algoritm, dimana initialization ini akan melakukan penentuan K-cluster secara otomatis berdasarkan datapoint yang diproses agar diperoleh detection rate yang tinggi dan false positif rate yang rendah.
- Serta menerapkan teknik windowing yaitu Sliding Window untuk mendapatkan efficient proses dan detection yang lebih baik.

2. Related Work

DDoS merupakan salah satu dari serangan pada jaringan internet yang mengakibatkan anomaly traffic. Beberapa teknik sudah dilakukan dalam pendeteksian serangan DDoS. Salah satunya dalam penelitian [4] melakukan teknik memonitor peningkatak jumlah IP Source yang masuk kedalam sebuah jaringan. Dalam penelitian ini melakukan scheme yang disebut Source IP address Monitoring (SIM) untuk mendeteksi Highly Distributed Denial of Service (HDDoS) attacks. Dalam scheme tersebut dilakukan pertama kali dengan melakukan off-line training dengan menyimpan IP address ke dalam IP Address Dataset (IAD) dengan melakukan penghapusan jika IP Address yang disimpan sudah expired. Kemudian dilakukan teknik detection and learning dengan melakukan menggambil beberapa data statistic dari interval waktu yang ditentukan. Dengan membandingkan IP Address baru yang muncul dalam interval waktu tersebut dengan IP Address yang berada pada IAD, dapat dihitung jumlah kemunculan IP Address yang baru dalam interval waktu tersebut. Jika rate per IP Address melebihi threshold, maka alarm akan berbunyi dan diperkirakan ini adalah sebuah serangan terhadap bandwidth.

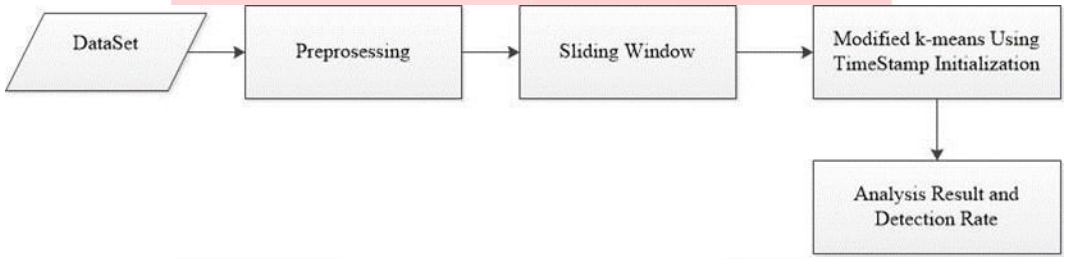
Pada [3] sudah mulai menerapkan k-means algoritm sebagai algoritma yang digunakan untuk melakukan deteksi dengan training data yang didifinisikan kedalam cluster dengan interval waktu antara normal dan anomaly. Dalam paper tersebut menyebutkan bahwa penentuan atau perbedaan K-cluster yang ditentukan akan memperngaruhi optimalitas dari detection. Pada penelitian ini juga terdeteksi adanya outlier yang mengakibatkan detection rate berkurang. Penentuan K-cluster sangat mempengaruhi tinggi kecilnya suatu deteksi jika menggunakan k-means algoritm.

Pada penelitian [6] melakukan penentuan k-cluster menggunakan improved k-means dimana teknik ini mirip dengan metode yang dipurpose pada tugas akhir ini. Pembeda antara penelitian [6] dengan tugas akhir ini, pertama terletak pada iterasi untuk menentukan centroid baru, improved k-means menggunakan optimalisasi dari Sum of Squared Deviations (SSD) sedangkan modified k-means with TimeStamp Initialization menggunakan rata-rata/mean($\bar{x}$) anggota suatu cluster. Kedua, improved k-means dalam hal penghitungan jarak antara datapoint dengan centroid menggunakan rumus informasi entropy dari attribut ditambahkan dengan rumus Euclidian distance, sedangkan modified k-means with TimeStamp Initialization menggunakan rumus Euclidian Distance. Ketiga, pada improved k-means dan modified k-means with TimeStamp Initialization sama-sama menggunakan Threshold sebagai acuan suatu data berada diluar suatu cluster, perbedaannya terletak pada parameter yang digunakan untuk menentukan Threshold tersebut. Pada improved k-means dijelaskan hanya menggunakan sebuah threshold statis untuk semua data percobaan, sedangkan modified k-means with TimeStamp Initialization dalam penentuan threshold berdasarkan parameter Sum of Squared Between terkecil yang dihasilkan dari beberapa pengujian treshold yang dilakukan oleh algoritma. Untuk alur percobaan, pada penelitian [6] menggunakan 2 dataset (dataset X, dataset Y) dimana dataset X digunakan sebagai dataset training untuk mendapatkan jumlah cluster yang digunakan, lalu diimplementasikan kedalam dataset Y. Pada tugas akhir ini, tidak menggunakan dataset training tetapi menggunakan Timestamp Initialization dan Sum of Squared Between terkecil sebagai acuan penentuan cluster, centroid dan threshold sehingga threshold yang digunakan dapat berubah berdasarkan dataset yang digunakan.

Penentuna K-cluster merupakan kunci untuk menghasilkan detection rate dan false positif rate yang tinggi maupun rendah. Pada paper ini kami akan memanfaatkan dan mempurpose TimeStamp Initialization dalam hal penentuan K-cluster dimana K-cluster ditentukan atau terbentuk berdasarkan datapoint yang masuk dari segi waktu pengerjaan / pemprosesan data dan mengoptimalkan detection rate dari k-means algoritm. Kami juga menggunakan teknik Sliding Window untuk mendapatkan detection yang lebih efficient.

3. METHODOLOGY

Pada bagian ini kami akan menjelaskan alur kerja dari proses kerja sistem detection yang kami purpose. Dataset yang kami gunakan adalah dataset KDDCup'99 dimana dataset ini akan kami gunakan sebagai bahan acuan dan datset testing dari penelitian kami. Sliding window kami gunakan sebagai pembatas dataset yang akan diolah, untuk efficient proses. Lalu, masuk kedalam modifikasi k-means menggunakan TimeStamp Initialization sebagai algoritma pendetectian dimana pada penelitian ini kami juga melakukan reset tentang threshold yang menjadi acuan sebuah data masuk ke dalam sebuah cluster tertentu atau tidak. Terakhir kami akan menarik kesimpulan dari dataset yang telah diproses oleh algoritma tersebut dari segi detection rate yang dihasilkan yang berdasarkan penentuan threshold yang kami deklarasikan. Untuk flow of proses seperti dibawah ini:



Gambar 3.1 Alur kerja sistem

3.1 Dataset

Dataset merupakan data sebagai dasar acuan dan data reset awal yang digunakan pada penelitian ini. Dataset yang digunakan diantaranya KDDCup'99 [7], Darpa'98 [8] dan dataset FlashCrowd. Untuk KDDCup'99 diambil 9 feature yaitu feature 23-31. Darpa'98 pertama dilakukan proses preprocessing dan dijadikan 9 feature. Untuk dataset Flash Crowd digunakan penggabungan dataset KDDCup'98 dan dataset Worldcup.

3.2 Sliding Window

Windowing merupakan teknik pembatasan data dengan melakukan pergeseran dan pengolahan data berdasarkan time-based atau count-based [7]. Dalam penelitian ini kami menggunakan Sliding Window dimana teknik ini melakukan pergeseran window secara pertahap,dimana pendeklarasian range of window [8] yang ditentukan diawal. Teknik ini juga pernah dipurpose dalam penelitian [9] untuk mencegah missing data yang dipadukan dengan algoritma Fuzzy k-means. Pada penelitian ini kami akan memanfaatkan teknik Sliding Window untuk memaksimalkan proses algoritma sehingga mendapatkan detection rate yang lebih akurat yang dipadukan dengan TimeStamp Initialization pada modified k-means.

3.3 Preprocessing

Preprocessing dilakukan untuk menggubah data mentah (capture traffic) Darpa dan WorldCup menjadi data baru yang terdiri dari 9 feature. Preprocessing menghasilkan dataset baru berupa feature unik dari capture traffic yang dilakukan, dari preprocessing dilakukan pendeteksian traffic menggunakan algoritma modified k-means menggunakan Timestamp initialization. Preprocessing dilakukan berdasarkan IP_Destination dan windowing selama 1 second. Untuk definisi setiap feature dapat dilihat seperti Tabel 3.0.2 Penjelasan feature preprocessing:

Tabel 3.1 Preprocessing

Count	Jumlah service yang terjadi
IP_Source	Jumlah IP Source yang meminta request
Protocol	Jumlah Protocol
SYN	Jumlah SYN
ACK	Jumlah ACK
Port_out	Jumlah Port_out
Length	Jumlah Length dengan panjang yang sama
Different_Source	Jumlah Different_Source
New_IP	Jumlah New_IP

3.4 Modified K-means

Modifikasi k-means yang dilakukan pada Tugas Akhir ini terletak pada penentuan jumlah k-cluster yang digunakan, jadi pada prosesnya penentuan k-cluster dilakukan oleh sistem berdasarkan data dan Threshold yang digunakan. Untuk alur proses algoritma modified k-means menggunakan metode Initialisasi cluster dan threshold dibawah ini:

3.4.1 Threshold

Threshold digunakan untuk menjadi batas atas suatu cluster dari centroid cluster tersebut. Setiap cluster memiliki threshold yang sama untuk acuan data, apakah data tersebut menjadi bagian suatu cluster atau tidak. Sistem menggunakan beberapa buah perbandingan threshold, yang diplih menggunakan Sum of Squared Between(SSB) terkecil dari beberapa threshold yang dicoba. Sum of Squared Between atau SSB adalah salah satu rumus perhitungan Sum of Squared dari Analysis of Variance (Anova) yang melakukan perhitungan berdasarkan pengurangan kuadrat mean dari data suatu class dengan mean of means dari total data dari semua class yang terbentuk. Rumus dari Sum of Squared Between:

Dimana SSB merupakan Sum of Squared Between, Xi adalah rata-rata dari setiap class (cluster) yang terbentuk, X adalah rata-rata dari jumlah setiap data dari semua class (cluster). Data atau nilai dari setiap class berupa jarak antara data tersebut terhadap centroid pada cluster. SSB menjadi acuan threshold yang baik untuk digunakan untuk data yang diproses pada window tertentu. Sehingga, Alur penentuan threshold mengikuti step dibawah ini:

1. pertama, inputkan threshold acuan awal.
2. Dari treshold awal dibentuk 50 threshold baru yang digunakan sebagai acuan batas atas cluster.
3. Testing data menggunakan semua threshold dan hitung SSB setiap threshold. Untuk rumus SSB seperti berikut:

$$SSB = \sum_{i=1}^{N} (\bar{X}_i - \bar{X})^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

4. Bandingan semua SSB threshold yang digunakan, dan ambil SSB terkecil. SSB terkecil dari sebuah threshold menjadi patokan threshold yang dipakai untuk memproses data pada window tersebut.
5. Gunakan threshold terpilih untuk menjadi acuan awal threshold pada window berikutnya.

3.4.2 Timestamp Initialization

Timestamp Initialization merupakan teknik yang digunakan untuk mendefinisikan jumlah k-cluster yang digunakan berdasarkan Threshold dan waktu paket diproses. Teknik ini memanfaatkan kedatangan paket sebagai acuan penentuan k-cluster dan centroid yang dikombinasikan dengan penggunaan Threshold sebagai batas atas dari sebuah cluster. Sehingga, bila suatu data berada diluar Threshold dari cluster yang sudah terbentuk maka data tersebut membentuk cluster baru dan menjadi centroid dari cluster tersebut. Jarak antara data dan centroid dihitung menggunakan rumus Euclidian Distance. Alur dari Timestamp Initialization mengikuti algorithma dibawah ini:

1. Data point pertama selalu dijadikan centroid pertama
2. Next data point akan diperiksa apakah termasuk kedalam sebuah centroid atau tidak. Proses ini menggunakan Treshold dan Euclidian distance sebagai rumus penghitung jarak antara centroid dengan datapoint; untuk rumus Euclidian distance seperti berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^{N}(x_i - y_i)^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

3. Iteration all cluster and data point to get a new centroid for all cluster has been created;
4. If there have another data point, repeat step 2 till 3 or not stop the proses;

4. ANALYSIS

Analisis dilakukan dengan menggunakan beberapa dataset berupa KDDCup'99, Darpa'98 week 3 Tuesday (Full Normal Traffic), Darpa'98 week 3 Wednesday (Normal Traffic + Attack), dan dataset FlashCrowd. Untuk analisis mengikuti:

1. Windowing yang digunakan Sliding Window dengan range data adalah 25.
2. Parameter hasil pengujian berupa TN, TP, FN, FP, DR, FPR, dan Accuracy yang dihasilkan oleh algoritma Modified K-means menggunakan Timestamp Initialization.
3. Analisis pengaruh SSB terhadap Threshold dan Accuracy yang dihasilkan oleh algoritma.

Parameter yang digunakan sebagai pengukur tingkat detection rate dan false positif rate yang diproleh adalah true positif, true negative, false positif, false negative yang akan dikalkulasi menghasilkan detection rate dan false positif rate yang dihasilkan oleh algoritma. Untuk penjelasan parameter dan rumus detection rate, false positive rate dan accuracy seperti berikut:

Tabel 4.1 Confusion matrix

		PREDICTION	
		ATTACK	NORMAL
ACTUAL	ATTACK	TP	FN
	NORMAL	FP	TN

$$Detection\ Rate = \frac{True\ Positive}{(True\ Positive + False\ Negative)} \quad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

$$False\ Positive\ Rate = \frac{False\ Positive}{(False\ Positive + True\ Negative)} \quad \ldots\ldots\ldots\ldots (4)$$

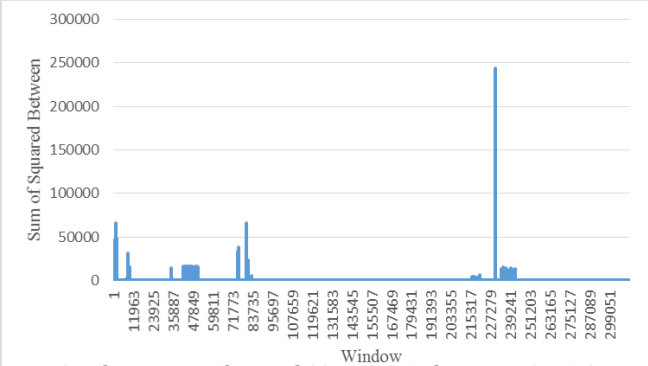
$$Accuracy = \frac{(TP + TN)}{(TP+TN+FP+FN)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Setelah melakukan pengujian dengan dataset yang telah ditentukan, Diperoleh hasil seperti dibawah ini:

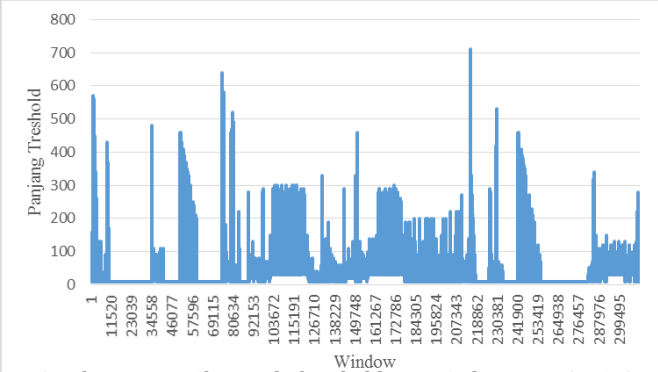
Tabel 4.2 hasil detection untuk beberapa dataset

Dataset	Rata-Rata TN	Rata-Rata TP	Rata-Rata FN	Rata-Rata FP	DR	FPR	Accuracy
KDDCup'99	16.31%	77.35%	3.17%	3.17%	96.06%	16.28%	93.65%
Darpa'98 Week 3 Tuesday	99.51%	0.00%	0.00%	0.49%	0.00%	0.49%	99.51%
Darpa'98 Week 1 Wednesday	94.59%	4.41%	0.29%	0.71%	86.13%	0.75%	99.00%
Flash Crowd per flow 2s	100.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%

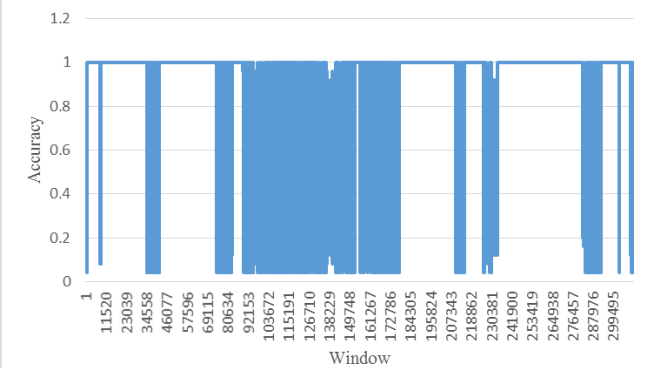
Dibawah ini merupakan hasil record dari penggunaan Threshold, SBB, Jumlah Cluster dan Accuracy per window yang dihasilkan dari pengujian dataset KDDCup'99.



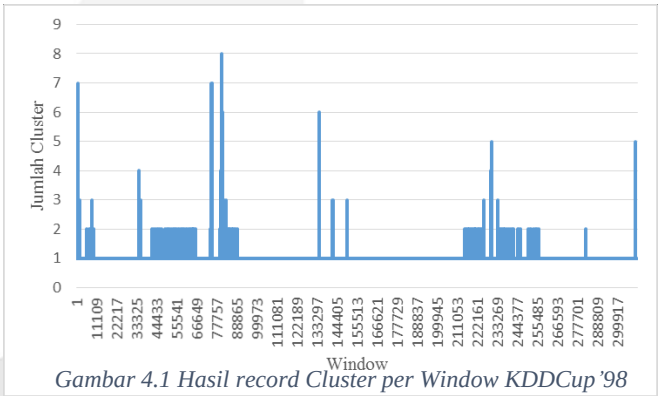
Gambar 4.4 Hasil record SSB per Window KDDCup'98



Gambar 4.3 Hasil record Threshold per Window KDDCup'98



Gambar 4.2 Hasil record Accuracy per Window KDDCup'98



Gambar 4.1 Hasil record Cluster per Window KDDCup'98

Dilihat dari gambar menyatakan bahwa pengaruh dari Sum of Squared Between yang dipilih mempengaruhi dari Threshold yang digunakan pada setiap window. Pemilihan Threshold yang digunakan berdasarkan SSB terkecil yang dihasilkan berdasarkan percobaan beberapa Threshold yang dilakukan algoritma modified k-means menggunakan Timestamp Initialization. Pemilihan Threshold mempengaruhi jumlah cluster yang dibentuk, jika SSB menghasilkan nilai 0 maka rata-rata cluster yang terbentuk adalah 1 buah cluster atau terbentuknya lebih dari 1 cluster tetapi dalam cluster tersebut hanya ada 1 buah anggota atau awalnya terdapat anggota dalam cluster tersebut tetapi setelah iterasi, mengakibatkan terjadinya kekosongan pada cluster tersebut. Hal tersebut yang mengakibatkan nilai SSB bernilai 0, karena mean of means dari semua cluster (jika terbentuk 1 cluster) yang dikurangkan dengan means setiap cluster akan menghasilkan nilai 0 (untuk cluster kosong tidak ikut dihitung meansnya). Untuk kecilnya accuracy yang dihasilkan pada suatu window, diakibatkan dari SSB yang bernilai 0 dan menghasilkan atau mengelompokkan data pada sebuah cluster dengan jarak data tidak melebihi threshold yang digunakan. Hal tersebut dikarenakan oleh data serangan maupun normal memiliki kemiripan yang sangat besar, yang mana pada sebuah cluster yang dikira normal terdapat serangan yang masuk pada cluster tersebut atau

sebaliknya. Hal ini yang mengakibatkan false positive rate yang dihasilkan pada percobaan dengan dataset KDDCup'99 cukup tinggi dan accuracy yang diperolehpun 93%.

5. CONCLUSION

Modified k-means menggunakan Timestamp Initialization dapat digunakan sebagai algoritma pengelompokan data traffic menggunakan 9 feature dengan similaritas tinggi untuk sebuah cluster dan similaritas rendah dalam setiap cluster yang terbentuk. Pemilihan Threshold mempengaruhi jumlah cluster yang dibentuk, jika SSB menghasilkan nilai 0 maka rata-rata cluster yang terbentuk adalah 1 buah cluster atau terbentuknya lebih dari 1 cluster tetapi dalam cluster tersebut hanya ada 1 buah anggota atau awalnya terdapat anggota dalam cluster tersebut tetapi setelah iterasi, mengakibatkan terjadinya kekosongan pada cluster tersebut. Pada pengujian dataset Darpa'98 week 3 Tuesday dimana dataset tersebut full normal diketahui terbentuknya cluster tidak selalu satu buah cluster, ini menandakan bahwa sebuah cluster normal tidak selalu dapat cluster kedalam sebuah cluster, ini dibuktikan dengan terjadinya False Positive Rate sebesar 0.49% pada pengujian menggunakan Dataset Darpa'98 week 3 Tuesday. Penentuan preprocessing yang tepat juga dapat meningkatan hasil deteksi yang dilakuakan, seperti pada pengujian dataset Darpa'98 dengan serangan dimana neptune dan smurf memiliki similaritas yang yang rendah dengan trafik normal sehingga menghasilkan accuracy yang tinggi dibandingkan dengan serangan lainnya (seperti back).

Untuk penelitian selanjutnya, dapat mencoba menggunakan parameter lain dalam hal penentuan Threshold. Dapat mencoba perhitungan Sum of Squared lain yang terdapat pada ANOVA (Analysis of Variance). Penggunaan preprocessing dan pembuatan feature yang tepat dapat memperbaiki akurasi yang dihasilkan, sehingga kedepannya bisa melakukan penelitian terhadap feature-feature pendukung lain (selain 9 feature yang ditampilkan pada Tugas Akhir ini) untuk mendeteksi serangan yang similaritasnya hampir serupa dengan trafik normal.

References

[1] V. L. L. Thing, M. Sloman and N. Dulay, "A Survey of Bots Used for Distributed Denial of Service Attacks," *IFIP International Federation for Information Processing*, vol. 232, 2007.

[2] 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.

[3] K. Garg and R. Chawla, "DETECTION OF DDOS ATTACKS USING DATA MINING," *International Journal of Computing and Business Research (IJCBR)*, vol. 2, no. 1, 2011.

[4] Tao Peng, Christopher Leckie, Kotagiri Ramamohanarao, "Proactively Detecting Distributed Denial of Service Attacks Using Source IP Address Monitoring," in *Networking 2004*, Springer Berlin Heidelberg, 2004, pp. 771-782.

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

[6] L. Liu, P. Wan, Y. wang and S. Liu, "Clustering and hybrid Genetic Algorithm based Intrusion Detection Strategy," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 12, pp. 762-770, 2014.

[7] I. University of California, "KDD Cup 1999 Data," [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html. [Accessed 25 05 2015].

[8] M. I. O. TECHNOLOGY, "LINCOLN LABORATORY," [Online]. Available: http://www.ll.mit.edu/ideval/data/. [Accessed 25 may 2015].

[9] P. S and O. P. Vyas, "Data Stream Mining: A Review on Windowing," *Global Journal of Computer Science and Technology Software & Data Engineering*, vol. 12, no. 11, 2012.

[10] Y. Iwakura, J. Suzuki, H. Yamada, Y. Yamaguchi, M. Tanabe and Y. Shoji, "An Efficient Sliding Window Processing for the Covariance Matrix Estimation," *IEEE 26th Convention of Electrical and Electronics Engineers in Israel (IEEEI)*, pp. 000977 - 000981, 2010.

[11] z. Liao, X. Lu, T. Yang and H. Wang, "Missing Data Imputation: AFuzzy K-means Clustering Algoritm over Sliding Window," *2009 Sixth International Conference on Fuzzy System and Knowledge Discovery*, vol. 3, pp. 133 - 137, 2009.