TEKNIK CLUSTERING PADA LOG FILE IDS BERDASARKAN TINGKAT SERANGAN MENGGUNAKAN ALGORITMA HIERARCHICAL CLUSTERING

[1]*Eriko Januar Akbar* (1210651284)
[2]*Triawan Adi Cahyanto, S.Kom, M.Kom*

Fakultas Teknik Program Studi Teknik Informatika
Universitas Muhammadiyah Jember
Email : erichokoko@gmail.com

ABSTRAK

Ada banyak survey yang dapat kita jadikan pedoman, di antaranya survey yang dilakukan oleh IDC (International Data Corporation) tahun 2003 yang mengatakan bahwa tantangan utama manajemen perusahaan dalam masalah keamanan sistem informasi adalah kurangnya sumber daya manusia(36%), disusul *monitoring* (19%), tidak adanya kebijakan dan prosedur (17%), serta integrasi antara solusi teknologi keamanan Sistem Informasi (12%).
Keamanan jaringan komputer sebagai bagian dari sebuah sistem sangat penting untuk menjaga validitas dan integritas data serta menjamin ketersediaan layanan bagi penggunaannya sistem keamanan jaringan komputer harus harus dilindungi dari segala macam serangan dan usaha-usaha penyusupan atau pemindaian oleh pihak yang tidak berhak. Algoritma Hierarchical clustering diterapkan pada data traning untuk menentukan banyaknya *cluster* yang diinginkan. Pada algoritma *clustering*, data akan dikelompokkan menjadi *cluster-cluster* berdasarkan kemiripan satu data dengan yang lain. Prinsip dari *clustering* adalah memaksimalkan kesamaan antar anggota satu *cluster* dan meminimumkan kesamaan antar anggota *cluster* yang berbeda. Selanjutnya dilakukan pelabelan yang akan mengetahui *active attak* dan *normal attack.* Output sistem digunakan untuk meng-*update* database.

Kata Kunci : *IDS, Hierarchical Clustering, Gene Cluster 3.0, Java Treeview.*

BAB I PENDAHULUAN

1.1. Latar Belakang

Perkembangan Teknologi Informasi (selanjutnya disebut TI) telah menyebabkan perubahan dan cara pandang hidup manusia dan suatu organisasi. Perkembangan TI yang sedemikian cepatnya telah membawa dunia memasuki era baru yang lebih cepat dari yang pernah dibayangkan sebelumnya. Ssat ini komputer tidak hanya berfungsi sebagai alat pengolahan data saja, namun telah menjadi senjata utama dalam berkompetisi.

Ada banyak survey yang dapat kita jadikan pedoman, di antaranya survey yang dilakukan oleh IDC (International Data Corporation) tahun 2003 yang mengatakan bahwa tantangan utama manajemen perusahaan dalam masalah keamanan sistem informasi adalah kurangnya sumber daya manusia(36%), disusul *monitoring* (19%), tidak adanya kebijakan dan prosedur (17%), serta integrasi antara solusi teknologi keamanan Sistem Informasi (12%)(Stiawan, 2005).

Keamanan jaringan komputer sebagai bagian dari sebuah sistem sangat penting untuk menjaga validitas dan integritas data serta menjamin ketersediaan layanan bagi penggunaannya sistem keamanan jaringan komputer harus harus dilindungi dari segala macam serangan dan usaha-usaha penyusupan atau pemindaian oleh pihak yang tidak berhak.

Algoritma Hierarchical clustering diterapkan pada data traning untuk menentukan banyaknya *cluster* yang diinginkan. Pada algoritma *clustering*, data akan dikelompokkan menjadi *cluster-cluster* berdasarkan kemiripan satu data dengan yang lain. Prinsip dari *clustering* adalah memaksimalkan kesamaan antar anggota satu *cluster* dan meminimumkan kesamaan antar anggota *cluster* yang berbeda. Selanjutnya dilakukan pelabelan yang akan mengetahui *active attak* dan *normal attack.* Output sistem digunakan untuk meng-*update* database rule SNORT.

Kemudian dapat dilakukan antisipasi agar serangan-serangan yang termasuk serangan aktif dan serangan normal setelah di*cluster* ada penangannya. Serangan normal akan dibuatkan rules baru sedangkan untuk serangan yang bersifat aktif akan dilakukan pencegahan.

Intrusion Detection System yang nantinya akan disebut IDS merupakan usaha mengindetifikasi adanya penyusup yang memasuki system tanpa otoritas (missal *craker*) atau seorang *user* yang sah tetapi menyalahgunakan *privilege* sumberdaya sistem. *Intrusion Detection System* (IDS) atau Sistem Deteksi Penyusupan adalah sistem komputer (bisa merupakan kombinasi *software* dan *hardware*) yang berusaha melakukan deteksi peyusupan IDS akan melakukan pemberitahuan saat mendeksi penyusupan IDS akan melakukan pemberitahuan saat mendeteksi sesuatu yang dianggap sebagai mencurigakan atau tindakan illegal. Salah satu program yang berfungsi sebagai IDS adalah snort**.**

1.2. Perumusan Masalah

Berdasarkan latar belakang masalah yang dijelaskan, dapat dirumuskan sebuah masalah yaitu :

1. Bagaimana implementasi Algoritma *Hierarchical Clustering* untuk mendeteksi serangan berdasarkan log file dari SNORT.
2. Bagaimana membagi tingkat serangan , yaitu *active attack* dan *normal attack.*

1.3. Batasan Masalah

Dalam pembuatan pengembangan intrusksi serangan dibuat beberapa batasan masalah agar permasalahan lebih terfokus. Adapun masalahnya sebagai berikut :

1. Sistem kinerja bisa melakukan clustering namun belum bisa melakukan tindakan aksi pencegahan.
2. Data set log file IDS diambil dari log file SNORT.
3. Serangan yang berhasil dideteksi berdasarkan data set yang digunakan yaitu :
   - *Port Scanning*
   - *Teardorp*
   - *Ip Spoofing*
   - *ICMP Flood*
   - *UDP Flood*
   - *Packet interception*
   - *Smurt attack*

1.4. Tujuan

Maksud dari penelitian ini adalah untuk pengembangan intruksi Snort menggunakan algoritma Hierarchical Clustering. Sedangkan tujuan yang diharapkan dari pengembangan intruksi snort ini adalah :

1. Mengetahui terhadap serangan-serangan aktif berdasarkan kinerja sistem.
2. Melakukan tindakan deteksi serangan aktif sehingga dapat membuat rules IDS yang baru

BAB II TINJAUAN PUSTAKA

2.1. *Intrusion Detection System* (IDS)

IDS adalah usaha mengindentifikasi adanya penyusup yang memasuki sistem tanpa otoritas (misal *cracker*) atau serangan user yang sah tetapi menyalahgunkan (*abuse*) sumberdaya sistem (Purbo, 2010).

Sistem Deteksi Intrusi atau IDS merupakan sebuah sistem yang mempunyai kemampuan untuk monitoring *traffic* jaringan, mendeteksi aktivitas-aktivitas yang mencurigakan, serta mampu melakukan pencegahan dini terhadap intrusi ataupun aktivitas-aktivitas yang dapat membahayakan sistem jaringan komputer yang telah dibangun. Beberapa keuntungan yang diperoleh dengan menggunakan IDS adalah sebagai berikut :

1. Sistem Deteksi Intruski dapat memberikan informasi apabila terdapat *worm* yang menyerang atau apabila sistem komputer telah berhasil dipenetrasi.
2. Sistem Deteksi Intrusi mampu memonitoring jaringan internal terhadap perilaku yang melanggar kebijakan sebuah institusi.
3. Sistem Deteksi Intrusi dapat membantu pelacakan setelah terjadinya suatu serangan, apa saj yang telah diserang dan dari mana serangan tersebut berasal.

Sitem Deteksi Intrusi mampu menyediakan informasi sejauh mana *firewall* bekerja dan siapa saja yang berhasil lolos melewatinya (Scott dkk, 2004).

2.2. Clustering

Clustering adalah proses mengempokkan atau penggolongan objek berdasarkan informasi yang di peroleh dari data yang menjelaskan hubungan antar objek dengan prinsip untuk memeasimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas/cluster. Clustering dalam data mining berguna untuk menemukan pola dstribusi di dalam sebuah data set yang berguna untuk proses analisa data. Kesamaan objek biasanya diperoleh dari kedekatan nilai-nilai atribut yang menjelaskan objek-objek data, sedangakan objek-objek data biasanya dipresentrasikan sebagai sebuah titik dalam ruang multi dimensi (Han, 2006).

2.3. Hierarchical Clustering

Algoritma *Hierarchical* dimulai dengan menjadikan objek menjadi sebuah *cluster* dan secara iterasi menggabungkan tiap *cluster* yang mirip. Jarak antara tiap obyek merupakan input dari algoritma ini. Iterasi terus berlanjut sampai semua obyek telah di*cluster* menjadi sebuah *cluster saja.* Output dari algoritma ini adalah sebuah *dendogram* (*Hierarchical tree*)(Edelstein, 1999).

2.4. Centroid Complete Linkage Hierarchical Method

Complete linkage memberikan kepastian bahwa semua item-item dalam satu cluster dalam jarak paling jauh (similaritas terkecil) satu sama lain.

Algoritma aglomerative pada umumnya dimulai dengan menentukan entri (elemen matriks) dalam D = $\{d_{ik}\}$ dan menggabungkan objek-objek yang bersesuaian misalnya *U* dan *V* untuk mendapatkan cluster (*UV*) untuk langkah (3) dari algoritma di atas jarak-jarak antara cluster (*UV*) dan cluster *W* yag lain dihitung dengan $d_{(UV)W}=maks\{d_{UW},d_{VW}\}$.

Di sini besar-besaran d_{UW} dan d_{VW} berturut-turut adalah jarak antara tetangga terdekat cluster-cluster *U* dan *W* dan juga cluster-cluster *V* dan *W* (Everitt:1993).

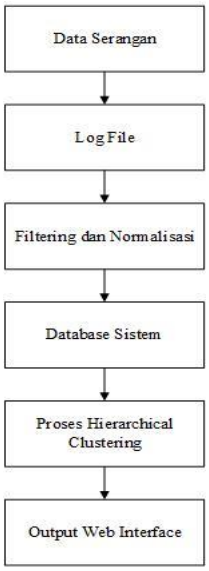
2.5. Basic security and Analysis engine (BASE)

BASE adalah sebuah *interface web* untuk melakukan analisis dari intrusi yang snort telah deteksi pada jaringan (Orebaugh, 2008:217) Base ditulis oleh Kevin Johnson adalah program analisi sistem jaringan berbasis PHP yang mencari dan memproses database dari *security event* yang dihasilkan oleh berbagai program monitoring jaringan, *firewall*, atau sensor IDS (Kohlenberg, 2007:424).

BAB III METODOLOGI PENELITIAN

3.1 Meteodologi

Data yang telah di download, selanjutnya di filtering parameter apa saja yang akan dipilih dan dimasukkan ke dalam database dan nantinya di analisa dan di clustering pada pre-processor dengan *hierarchical clustering*, lalu hasil di tampilkan pada user interface yang menggunakan aplikasi Gene Cluster dan Java Treeview untuk melihat hasil visualisasi. Output dari Java Treeview dapat terlihat yang termaksuhkserang aktif dan bukan serangan.



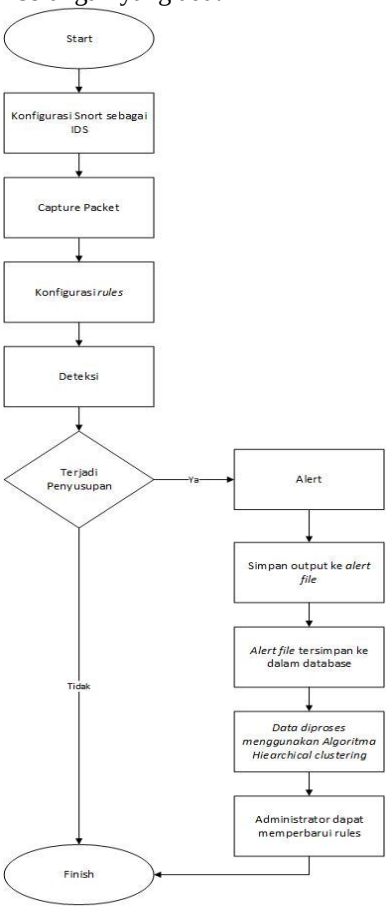
Gambar 3. 1 Desain Sistem

2.6. Sumber Data

Data yang digunakan pada penelitian ini merupakan data yang diambil dari log data serangan yang diambil dari correctedKDD 199 cup.

2.7. Perancangan Sistem

Sistem yang akan dibangun dalam penelitian ini adalah sistem berbasis web yang dirancang mempermudah webmin untuk mengantisipasi serangan-serangan yang ada.



Gambar 3.2. Flowcart Perancangan Sistem

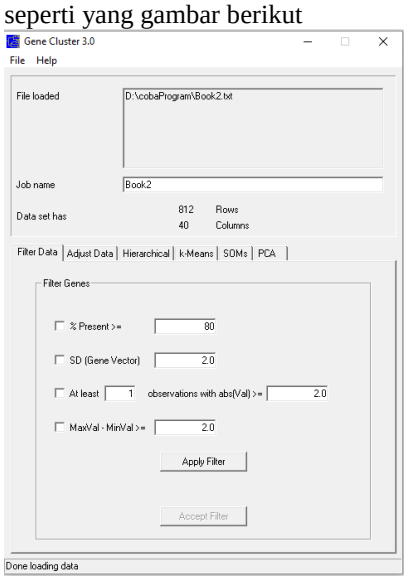
Flowcart yang ditunjukkan merupakan gambaran bagaimana kerja IDS secara keseluruhan. Pertama menkonfigurasi Snort sebagai IDS kemudian paket akan di*capture* lalu dideteksi IDS berdasarkan rule yang tersedia. Kemudian apabila tidak terdeteksi maka proses berakhir. Apabila serangan terdeteksi, maka IDS akan menyimpan alert ke log file dan mengirimkan alert pesan serangan ke *admistrator* yang akan di tampilkan pada *web interface* yang disimpan pada databse untuk diolah kembali menggunakan algoritma *Hierarchical Clustering* untuk mengenali jenis serangan *Network Package*. Analisa *Hierarchical Intrusion Detection Engine* (HIDE) dari aktifitas Hacking yang digolongkan sebagai *Active Attack* dan *Normal Attack.*

.

BAB IV IMPLEMENTASI DAN PENGUJIAN

4.1 Implementasi

Data yang akan diuji diupload ke Gene cluster 3.0 seperti yang gambar berikut

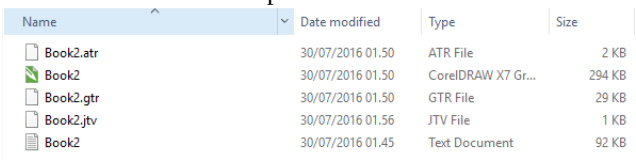


Gambar 4. 1 Pengujian Data Menggunakan Gene Cluster 3.0

Kemudian data akan diolah menggunakan hierarchical clustering dengan klik tombol hierarchical kemudian centang cluster, selanjutnya pilih *Euclidean Distance* untuk mengukur jarak antar cluster. Pilih tombol *Complete linkkage* untuk centroidnya seperti gambar berikut

Gambar 4. 2 Pengujian Gene Cluster Menggunakan *Hierarchical Clustering*

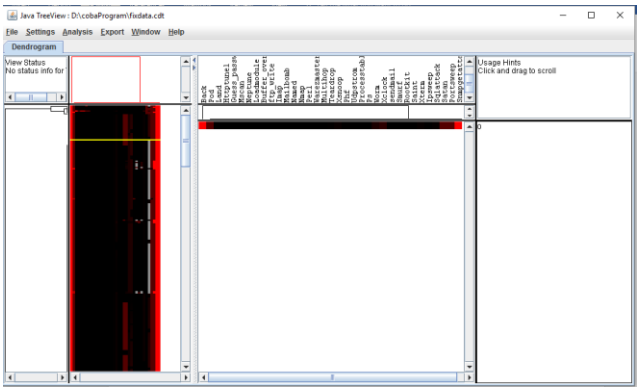
Setelah data diproses akan menghasilkan file berbentuk .atr .gtr .jtv dan .cdt. untuk menganalisa hasil dari *Hierarchical Clustering* menggunakan Java Treeview. Berikut contoh file hasil proses Gene tool 3.0



Gambar 4. 3 File Hasil Proses Cluster

4.2 Visualisai menggunakan Java Treeview

File yang akan diguanakan adalah format cdt akan bisa dibaca oleh Java Treeview. Berikut file hasil gambar Java Treeview setelah upload data



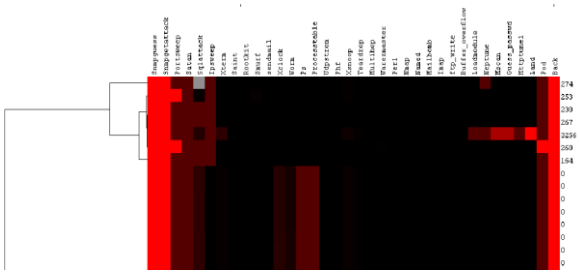
Gambar 4. 4Output hasil yang Diolah Java Treeview

Dari hasil gambar diatas bahwa :

1. Warna merah bernilai positif
2. Warna hitam bernilai 0
3. Warna hijau bernilai negatif
4. Warna abu bernilang hilang

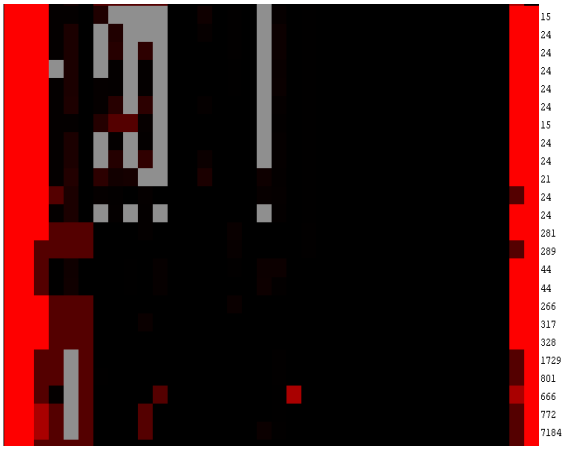
4.3 Hasil Analisa Java Treeview

Dalam melakukan analisa data KDD 1999 terdapat banyak data bernilai 0 sehingga hasil visualisasi data tersebut banyak mengandung warna hitam yang berarti bernilai 0.



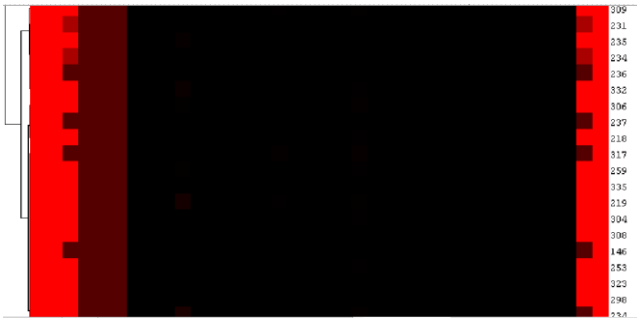
Gambar 4. 5 Hasil Visualisasi Algoritma *Hierarchical Clustering*

Gambar 4.5 memperlihatkan serangan yang terjadi kebanyakan pada *snmpguess* dan *snmpgeattack* mendominasikan sehingga berwana merah yaitu positif serangan. Serta diikuti *portsweep*, *satan*, *sql attack*, *ipsweep* dan *pod* berwarna merah. Sedangan port yang lain berwana hitam yang bernilai zero atau tidak terjadi serangan dianggap normal.



Gambar 4. 6 Hasil Visualisasi Algoritma *Hierarchical Clustering* (2)

Gambar 4.10 lanjutan dari gambar 4.9 menunujukan terdapat nilai yang *missing* menandakan nilai tersebut hilang atau tidak terhitung karena hasil dari nilai kedekatan jarak bernila min atau kurang dari 0. Dari gamabar 4.10 juga masih didominasi oleh *snmpguess*, *snmpgeattack* dan *back*. Serangan terus terjadi dan memiliki nilai sehingga tidak terlihat berwarna merah.



Gambar 4. 5 Hasil Visualisasi Algoritma *Hierarchical Clustering* (2)

Pada gambar ini memperlihatkan bahwa serangan yang bernilai positif yaitu berwarna merah, sedangkan nilai 0 mendominasi yang ditujukan oleh warna hitam. Terdapat mising data yaitu ditukan oleh warna abu2. Dari data yang telah di oleh tersebut tidak ada nilai negatif karena tidak ada warna hijau yang menunjukkan nilai negatif.

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan

Setelah Gene cluster 3.0 melakukan cluster dan Javatreeview melakukan visualisasi dapat disimpulkan bahwa :

1. Algoritma *Hierarchical Clustering* dapat mendeteksi serangan aktif dan serangan normal menggunakan Gene cluster 3.0 berdasarakan log file yang telah ada

2. Tingkatan serangan dapat dibagi dengan memanfaatkan metode *Hierarchical Clustering* mekanisme kerja melakukan clustering sesuai dengan kedekan data.
3. Telah dapat mengetahu terhadap serangan-serangan aktif berdasarakan kinerja sistem.

5.2 Saran

Melihat hasil analisis algoritma *Hierarchical Clustering* menggunakan Gene cluster 3.0 dan Java Treeview. Ada beberapa saran untuk penyempurnaan dalam peningkatan algortima *Hierarchical Clustering* ini :

1. Data yang digunakan dalam tugas akhir ini kurang bervariasi sehingga banyak hasil yang terdefinisikan bernilai 0 dan tidak ada hasil yang bernilai negatif. Sehingga data harus memiliki nilai yang bervariasi supaya dengan menggunakan metode *Hierarchical Clustering* dapat divisualisasikan dengan baik.
2. Untuk dilakukan penelitian lebih lanjut disarankan peneliti selanjutnya akan membuat program untuk mengolah log file yang banyak jumlahnya supaya hasil yang didapat sesuai dengan yang diharapkan.

DAFTAR PUSTAKA

Han, J. 2006. *Data Mining Concepst and Techniques Second Edition*, Bloomington, USA.

Edelstein, H. A. 1999. *Introduction to Data Minning and Konwledge Discovery*, Third Edition. Two Crows Corporation, USA.

Pinkard, B. dan Orebaugh, A. 2008. *Nmap in the Enterprise : Your Guide to Network Scanning*, syngress, United state of America.

Endorf, C., Schultz, E., dan Mellander, J. 2004. *Intrusion Detection & Prevention.* Emeryville, California.

Han, J., dan Micheline. 2001. *Data Minning Concepts and Technique*, Morgan Kaufman, USA.

Margo, R, dan Eka R. K. 2014. *Penerapan Algoritma Clustering untuk Mengelompokkan Ketertarikan Siswa Berdasarkan Aktivitas di Mode Pembelajaran Elektronik*. STMIK AMIKOM, Yogyakarta.

Stiawan, D. 2005. *Sistem Keamanan Komputer*, PT Elex Media Komputindo, Jakarta.

Syafrizal, M. 2005. *Pengantar Jaringan Komputer*, Penerbit ANDI, Yogyakarta.

Everitt, B.S. 1993. *Cluster Analysis* (3ed). Edward Arnold, London.

Scott, C., Wolfe P., dan Hayes B., 2004. *SNORT for Dummies*, Willey Publising Inc, USA.

Ariyus, D. 2007. *Intrusion Detection System*, Penerbit ANDI, Yogyakarta.

Raifudin, R. 2010. *Mengayang Hacker dengan SNORT*. Penerbit ANDI, Yogyakarta.

Anh, L. 2008, *On Optimizing Load Balancing of Intrusion Prevention and Prevention System*, IEEE, INFOCOM workshops.