1. Pendahuluan

1.1 Latar belakang

Data dan informasi sensitif dalam skala besar diproses dalam jaringan komputer. Sehingga diperlukan suatu sistem keamanan dalam jaringan komputer yang tahan dan toleran terhadap intrusi jaringan. Intrusi jaringan merupakan upaya untuk mendapatkan akses ilegal ke *resource* jaringan atau mem-*by-pass* sistem keamanan yang ada. Upaya itu dapat dilakukan dengan mengakses sistem melalui Internet, maupun melalui celah keamanan jaringan lainnya. Oleh karena itu, *Intrusion Detection System* (IDS) diperlukan untuk mengatasi permasalahan tersebut. *Intrusion detection* adalah suatu proses *monitoring* kejadian yang terjadi pada sistem komputer atau jaringan serta menganalisisnya untuk mengetahui aktivitas tersebut termasuk normal atau intrusi.

Dalam model proses IDS terdapat tiga fungsi dasar. Pertama, pengambilan data dari berbagai *level* dari sistem seperti *network, host,*dan *application.* Kedua, analisis data yang diperoleh untuk mengenali intrusi. Pendekatan yang sering digunakan untuk mengenali intrusi yaitu *anomaly detection* dan *misuse detection/signature analysis*. Ketiga, respon terhadap serangan. Ada dua macam: respon aktif dalam hal ini berarti melakukan beberapa aksi secara otomatis untuk mengintervensi sistem yang ada, sedangkan pasif adalah memberikan *report* pada administrator yang akan melakukan respon terhadap sistem.

Anomaly detection dapat mendeteksi serangan dalam host atau network yang menyimpang dari aktivitas normal berdasarkan probabilitas statistika. *Statistical anomaly detection* tidak memiliki model *intelligent learning* yang mungkin menyebabkan *false alarm* memiliki tingkat deteksi tinggi. *Anomaly detection* menggunakan pendekatan *unsupervised learning(*tidak ada *intelligent learning)*, yang mampu mendeteksi intrusi tanpa harus mempelajari data sebelumnya[9]. *Misuse detection/signature analysis* membutuhkan akses ke database besar dari *signature intrusion* yang diketahui. Detektor melakukan analisis terhadap aktivitas sistem, mencari *event* atau set *event* yang cocok dengan pola perilaku yang dikenali sebagai serangan.

Dalam tugas akhir ini digunakan pendekatan metode *anomaly detection*. *Anomaly detection* yang merupakan salah satu tugas dari proses data *mining*, diharapkan dapat menemukan objek yang berbeda dari kebanyakan objek yang ada. Seringkali objek anomali tersebut dikatakan sebagai *outlier* atau pencilan. Metode deteksi anomali dapat mendeteksi serangan berdasarkan statistik probabilitas, yang memungkinkan untuk generalisasi dan membantu dalam pendeteksian instrusi baru. Namun, statistik deteksi anomali tidak didasarkan pada *adaptive intelligent model* dan tidak bisa dipelajari dari pola aktivitas normal maupun *malicious traffic pattern (unsupervised learning)* [5].

Bayesian Network (BN) adalah representasi grafis dari gabungan dari probabilitas fungsi distribusi lebih dari satu set variabel. Struktur jaringan direpresentasikan sebagai *Directed Acyclic Graph* (DAG) di mana setiap *node* berkorespondensi dengan variabel acak dan setiap *edge* menunjukkan hubungan ketergantunga antar variabel-variabel yang terhubung [5]. Setiap variabel memiliki satu finite set *mutually exclusive states.*

Dalam Tugas Akhir ini metode *Bayesian Network* yang digunakan yaitu *Tree Augmented Naive Bayes* (TAN) *classifier*. Klasifikasi dengan TAN memiliki beberapa kelebihan dalam pemodelan data. Pertama, TAN memodelkan semua *dependencies* antar variabel, tujuannya agar mudah dalam menangani kasus beberapa entri data yang hilang. Kedua, TAN dapat digunakan untuk mempelajari hubungan *causal*, sehingga dapat digunakan untuk memperoleh pemahaman tentang domain masalah dan memprediksi konsekuensi dari intervensi. Ketiga, karena dimodelkan dengan *causal* dan semantik probabilistic maka akan membentuk representasi ideal untuk menggabungkan *prior knowledge* (yang sering datang dalam bentuk *causal*) dan data. Metode statistik Bayesian dan *Bayesian Network* merupakan pendekatan yang efisien dan berprinsip untuk menghindari overfitting data [14]. Sistem Bayesian mempunyai dasar matematika yang kuat dalam menangani implementasi IDS. *Bayesian Network* IDS harus membedakan antara intrusi/serangan dan aktivitas normal jaringan dengan membandingkan *metric* dari setiap *network traffic sample*.

Adaptive network IDS dengan medote BN mengambil data *offline* dari dataset KDD Cup tahun 1999 untuk mengukur kelayakan dan efektivitas sistem. Tipe datanya ada kontinu dan nominal dengan mengambil 9 dari 41 *feature*s termasuk *unbalanced* data[12]. Karena setiap variabel (*node*) dalam BN dikaitkan dengan *Conditional Probability Tabel* (CPT), yang menyebutkan probabilitas bersyarat untuk variabel yang memberikan semua kombinasi induknya. Akurasi deteksi intrusi dapat dilihat dari seberapa dekat kah data klasifikasi dengan data asli. Dengan BN sebagai model yang ideal untuk menggabungkan *prior knowledge* sebelumnya dengan data baru dan menyimpulkan menjadi *posterior knowledge*. Oleh karena itu, *Bayesian Network* dengan TAN *classifier* diharapkan dapat meningkatkan tingkat akurasi IDS.

1.2 Perumusan masalah

Dengan mengacu pada latar belakang masalah, maka permasalahan yang dibahas dan diteliti adalah :

1. Bagaimana mengimplementasikan metode *Bayesian Network* untuk deteksi anomali pada IDS.
2. Bagaimana mengevaluasi performansi dari metode *Bayesian Network* untuk mendeteksi terjadinya anomali dengan kasus data intrusi.

1.3 Batasan masalah

Batasan masalah yang digunakan dalam penelitian ini antara lain:

1. Pendekatan yang digunakan adalah pendekatan analisis deteksi anomali, dengan menggunakan metode *Bayesian Network* untuk menganalisa.
2. Data yang digunakan adalah *network connection record*.
3. Menggunakan data *offline,* yaitu data yang digunakan pada KDD Cup tahun 1999.
4. Menggunakan tools Microsoft Excel untuk *preprocessing* data.
5. Tidak dapat digunakan untuk mendeteksi intrusi yang dilakukan oleh pihak yang mempunyai akses dalam sistem komputer tersebut, dan intrusi yang dilakukan dengan menyerupai data normal.
6. Data set yang digunakan untuk dianalisis tidak boleh didominasi oleh intrusi.

1.4 Tujuan

Berdasarkan rumusan masalah, maka tujuan dari tugas akhir ini adalah:

1. Mengimplementasikan metode *Bayesian Network* dengan TAN *classifier* untuk mendeteksi anomali yang terjadi dalam suatu jaringan.
2. Menganalisis performansi deteksi anomali dengan metode *Bayesian Network* berdasarkan prediksi nomal (*true negative rate* dan *false negative rate)* dan prediksi intrusi *(true positive rate* dan *false positive rate).*

1.5 Metodologi penyelesaian masalah

Metode yang digunakan dalam penyelesaian tugas akhir ini adalah menggunakan metode studi pustaka atau studi literatur dan analisis dengan langkah kerja sebagai berikut :

1. Studi Literatur :
   a. Pencarian referensi, dan sumber-sumber lain yang berhubungan dengan data *mining*, *Bayesian Network*, *TAN classifier*, *network connection record*, *intrusion*, jaringan komputer dan pengukuran evaluasi pada IDS(*Intrusion Detection System*).
   b. Pendalaman materi, mempelajari dan memahami materi yang berhubungan dengan tugas akhir.
2. Melakukan analisis dan desain perangkat lunak yang dibangun. Perancangan akan dimulai dengan membangun skema / *flowchart* tentang alur sistem keseluruhan yang akan dibangun. Dilanjutkan dengan mencari kebutuhan-kebutuhan yang diperlukan oleh perangkat lunak dan sistem. Kemudian, mempersiapkan data yang akan diolah termasuk data *prepocessing* didalamnya.
3. Melakukan implementasi perangkat lunak mengunakan Matlab R2008a.
4. Melakukan analisis performansi dari hasil pengujian perangkat lunak yang telah diciptakan dengan melihat tingkat *detection rate* dan *false positive rate*.
5. Pengambilan kesimpulan dan penyusunan laporan Tugas Akhir.