ANALISA DAN IMPLEMENTASI METODE HIDDEN MARKOV MODEL PADA INTRUSION DETECTION SYSTEM (IDS)

Manaek Yudibert Donny Pasaribu[1], Jondri[2], Angelina Prima Kurniati[3]

[1]Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Penerapan data mining pada Intrusion Detection System didasari oleh kemampuan data mining dalam mengekstrak informasi-informasi berharga dari sekumpulan besar data. Hal ini bersesuaian dengan banyaknya limpahan data yang terdapat dalam masalah Intrusion Detection System, sehingga kemudian dapat diketeahui pola-pola tersembunyi yang dapat memberikan informasi yang akurat tentang adanya intrusi atau serangan.

Pada tugas akhir ini telah diimplementasikan penggunaan metode Hidden Markov Model dalam pembangunan misuse Detection model yang merupakan salah satu pendekatan dalam membangun Intrusion Detection System. Hidden Markov Model merupakan algoritma mengimplementasikan fungsionalitas klasifikasi dengan menggunakan pendekatan statistika dalam memprediksi keanggotaan suatu kelas. Keberhasilan metode Hidden Markov Model yang pernah diterapkan dalam pengenalan suara merupakan ide awal dari pembuatan tugas akhir ini.

Kata Kunci : Hidden Markov Model, Intrusion Detection System, misuse Detection

Abstract

Data mining implementation in Intrusion Detection System based on the data mining ability to extract the important information from large dataset. This issue correlated with so many data in Intrusion Detection System problem, then the hidden pattern that give the accurate information about Intrusion or attack can be known.

In this final project has been implemented the usage of Hidden Markov Model method to develop the misuse Detection model which is one of approaching method to develop Intrusion Detection System. Hidden Markov Model is the algorithm that implemented the classification functionality by using statistic approaching to predict member in the class. This final project inspired by the successful of Hidden Markov Model in speech recognition.

Keywords : Hidden Markov Model, Intrusion Detection System, misuse Detection

1. PENDAHULUAN

1.1 Latar belakang

Jaringan komputer terus mengalami perkembangan, baik dari skalabilitas, jumlah node dan teknologi yang digunakan. Dengan begitu jaringan komputer memerlukan pengelolaan yang baik agar ketersediaan jaringan selalu tinggi. Dalam hal ini keamanan jaringan komputer merupakan salah satu hal yang perlu diperhatikan secara khusus.

Penyusupan ataupun yang sering disebut dengan intrusi adalah salah satu jenis gangguan jaringan komputer. Definisi intrusi adalah sesuatu yang berusaha merusak atau menyalahgunakan sistem, atau setiap usaha yang melakukan *compromise* integritas,kepercayaan atau ketersediaan suatu sumberdaya komputer [2]. Intrusi berpotensi mengakibatkan ancaman terhadap integritas jaringan tanpa diketahui oleh pemiliknya. *Intrusion Detection System* (IDS) merupakan *tools* yang sangat populer saat ini yang digunakan untuk mengidentifikasi adanya intrusi didalam suatu jaringan[8].

Salah satu pondasi *tools* IDS adalah kemampuannya dalam memprediksi atau menggolongkan data hasil audit sebagai sebuah serangan atau bukan. Dalam implementasinya banyaknya jumlah data audit yang akan diprediksi, cepatnya pertambahan variasi tipe serangan atau intrusi membutuhkan mekanisme yang bekerja secara cerdas dan efektif. Dalam konteks inilah *data mining* diperlukan sehubungan dengan kemampuannya untuk mendapatkan informasi yang berharga dari sekumpulan besar data.

Banyak sekali metode statistika yang dapat digunakan untuk melakukan prediksi didalam *data mining*. Salah satunya adalah *Hidden Markov Model* (HMM). *Hidden Markov Model* merupakan suatu model statistik yang dapat digunakan untuk menganalisa data sekuensial. HMM akan menggunakan parameter-parameter untuk menghasilkan suatu model. HMM menggunakan kombinasi linear dari setiap kemungkinan dan menggunakan pendekatan yang berdasar frekuensi dari kejadian dan nilai prediktif. Sudah banyak studi kasus yang dapat diselesaikan dengan metode HMM ini. Diantaranya adalah *Speech Recognition* dan pada bioinformatika yaitu mengimplementasikan pencarian gen pada data sekuens struktur genome untuk beberapa organisme.

Dalam tugas akhir ini akan dilakukan analisa dan implementasi teknik *data mining* pada domain *Intrusion Detection System (IDS)*. Metode HMM akan dipakai untuk menentukan karakteristik akses yang digolongkan kedalam suatu intrusi atau normal dengan menggunakan atribut *Network based*.

1.2 Perumusan masalah

Dalam tugas akhir ini, penulis akan mencoba untuk mengimplementasikan metode *Hidden Markov Model* pada pembangunan suatu *Intrusion Detection System* dan mencoba untuk menarik kesimpulan dari implementasi tersebut. Adapun rumusan masalahnya adalah sebagai berikut:

a. Apakah metode *Hidden Markov Model* dapat digunakan dalam studi kasus *Intrusion Detection System* ?

b. Bagaimana cara pembuatan model dengan menggunakan metode *Hidden Markov Model* dengan menggunakan dataset KDD cup 1999 ?

c. Implementasi yang dilakukan hanya sampai terbentuknya misuse *Detection* model.

d. Data set yang digunakan sudah melalui proses diskritisasi dan feature selection.

e. Pengukuran keakuratan, sensitivitas pencarian intrusi dari metode *Hidden Markov Model.*

1.3 Tujuan

Tujuan dari tugas akhir ini adalah :

a. Mempelajari metode *Hidden Markov Model* dan pemanfaatannya pada *data mining.*

b. Menganalisa penggunaan *Hidden Markov Model* jika diterapkan dalam *Intrusion Detection*.

c. Mengukur tingkat efisiensi dan akurasi metode *Hidden Markov Model* pada *Intrusion Detection* dan melakukan implementasi dari proses.

1.4 Metodologi penyelesaian masalah

Metode yang akan digunakan untuk menyelesaikan tugas akhir ini adalah :

a. Studi Literatur

Mempelajari dasar teori dan literatur-literatur tentang *Data mining, Hidden Markov Model*, *Intrusion Detection, Data set* KDD cup 1999 dan mempelajari cara pembangunan aplikasi.

b. Perencanaan

Tahapan ini dilakukan untuk melakukan perencanaan tentang apa yang akan dikerjakan dan apa yang perlu dipersiapkan.

c. Perancangan Perangkat Lunak

Bertujuan untuk melakukan analisa dan perancangan pengembangan perangkat lunak dengan menggambarkan dalam modul-modul perangkat lunak.

d. Pembuatan Perangkat Lunak

Melakukan implementasi metode *Hidden Markov Model* pada perangkat lunak sesuai dengan analisa perancangan yang telah dilakukan.

e. Pengujian dan perbaikan perangkat lunak

Dalam tahap ini akan diuji program untuk berbagai training data set dan mencari kesalahan-kesalahan yang masih muncul dalam perangkat lunak.

f. Analisa terhadap hasil pengujian perangkat lunak.

g. Pengambilan kesimpulan dan penyusunan laporan.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari hasil pengujian, dapat disimpulkan beberapa hal sebagai berikut :

1. Berdasarkan hasil pengujian dapat disimpulkan bahwa penambahan jumlah *state* dan jumlah *observation symbol* tidak selalu menghasilkan model yang memiliki *Detection rate* yang lebih baik
2. Penggunaan metode inisialisasi *random* dapat digunakan karena tidak mempengaruhi model secara signifikan.
3. Penggunaan dataset *distinct* pada proses distinct dapat menaikan nilai *Detection rete* tetapi juga dapat menurunkan nilai akurasi.
4. Penggunaan metode Hidden Markov Model untuk mengenali pola data yang tidak ada pada dataset tidak terlalu baik.

5.2 Saran

1. Untuk menghasilkan model yang baik, diperlukan pertimbangan yang baik dalam hal penggunaan data uji yang berulang. Karena hal tersebut akan mempengaruhi nilai *Detection rate* dan nilai akurasi dari model yang diciptakan.
2. Melakukan pengujian terhadap data yang telah melalui proses yang menangani masalah *imbalance data.*
3. Mengimplementasikan metode *Hidden Markov Model* yang lebih kompleks seperti penggabungan metode *Hidden Markov Model* dengan algoritma genetika, pengaplikasian metode *hidden markov model* yang dapat menangani data uji yang *continues,* ataupun pengaplikasian metode *Hybrid Hidden Markov Model.*
4. Pada percobaan ini *preprocessing* tidak diperhatikan secara khusus. Tidak menutup kemungkinan untuk melihat pengaruh *preprocessing* terhadap model.

Referensi

[1] Rabiner, L. R. & Juang, B. H. 1986. "Introduction to Hidden Markov Models". IEEE ASP Magazine.

[2] Rabiner, L. R. 1989. "A tutorial on hidden Markov models and selected applications in speech recognition". Proceedings of the IEEE, 77, 2, 257-286

[3] Chen, Huiping. 2004. *Data mining* approaches for intrusion detection.

[4] Jecheva, Veselina. *About Some Applications of Hidden Markov Model in Intrusion Detection Systems*. International Conference on Computer Systems and Technologies - *CompSysTech'06*

[5] Shrijit S. Joshi and Vir V. Phoha. "Investigating Hidden Markov Models Capabilities in Anomaly Detection" Computer Science, Louisiana Tech University, 2005

[6] Resch, Barbara. *Hidden Markov Model.* Signal Processing and Speech Communication Laboratory Inffeldgasse, (February 2007)

[7] Noreen, Nita. 2004. "Penggunaan Hybrid HMM dan GA dalam Pengenalan Ucapan yang Tidak Bergantung Pembicara". Bandung: STT Telkom

[8] Han, Jiawai., Micheline Kamber. 2001. *Data mining* : Concepts and Techniques. Simon Fraser University : Morgan Kaufmann.

[9] Kuchimanchi, G., Phoha, V.V., Balagani,K.S. and Gaddam, S.R., imension Reduction using Feature Extraction Methods for Real-time Misuse Detection Systems. In W*orkshop on Information Assurance, United States Military Academy, West Point, NY*, (2004).

[10] KDD Data Set 1999. Http:// kdd.ics.uci.edu/ databases / kddcup99/ kddcup99.html

[11] Kumar, Vipin. 2004. *Data mining* for network intrusion detection. University of Minnesota