BAB II

LANDASAN TEORI

Pada BAB II ini akan disampaikan materi-materi yang berkaitan dengan *Intrusion Detection system* (IDS) dan algoritma *K-Nearest Neighbor* (K-NN), yang merupakan landasan bagi pembahasan *K-Nearest Neighbor* (K-NN) untuk aplikasi IDS.

2.1 Intrusion Detection system (IDS)

IDS merupakan sebuah aplikasi yang mampu mencatat kegiatan dalam suatu jaringan dan menganalisa paket-paket yang dikirim melalui lalu lintas jaringan secara *realtime*. Sistem IDS sendiri merupakan sistem sensor yang berfungsi sebagai memonitoring paket-paket yang dianggap mencurigakan dan menyimpan setiap *log* ke database.

Tujuan dari sistesm ini yaitu mengawasi jika terjadi penetrasi ke dalam sistem, mengawasi *traffic* yang terjadi pada jaringan, mendeteksi *anomaly* terjadinya penyimpangan dari sistem yang normal atau tingkah laku user, mendeteksi *signature* dan membedakan pola antara signature user dengan attacker (Alamsyah, 2011).

2.1.1 Cara Kerja IDS

IDS juga memiliki cara kerja dalam menganalisa apakah paket data yang dianggap sebagai *intrusion* oleh *intruser*. Cara kerja IDS terbagi menjadi dua, yaitu (Alamsyah, 2011):

a. Knowledge Based

Knowledge Based pada IDS adalah cara kerja IDS dengan mengenali adanya penyusupan dengan cara menyadap paket data kemudian membandingkannya dengan database *rule* pada IDS tersebut. Database *rule* tersebut dapat berisi *signature* paket serangan. Jika *pattern* atau pola paket data tersebut terdapat kesamaan dengan *rule database* pada IDS, maka paket data

tersebut dianggap sebagai serangan dan demikian juga sebaliknya, jika paket data tersebut tidak memiliki kesamaan dengan rule database pada IDS, maka paket data tidak dianggap serangan.

b. *Behavior Based*

Behavior based adalah cara kerja IDS dengan mendeteksi adanya penyusup dengan mengamati adanya kejanggalan pada sistem, atau adanya keanehan dan kejanggalan dari kondisi pada saat sistem normal, sebagai contoh adanya penggunakan memory yang melonjak secara terus menerus atau terdapatnya koneksi secara parallel dari satu IP dalam jumlah banyak dan dalam waktu yang bersamaan. Kondisi tersebut dianggap kejanggalan yang selanjutnya oleh IDS *anomaly based* dianggap sebagai serangan.

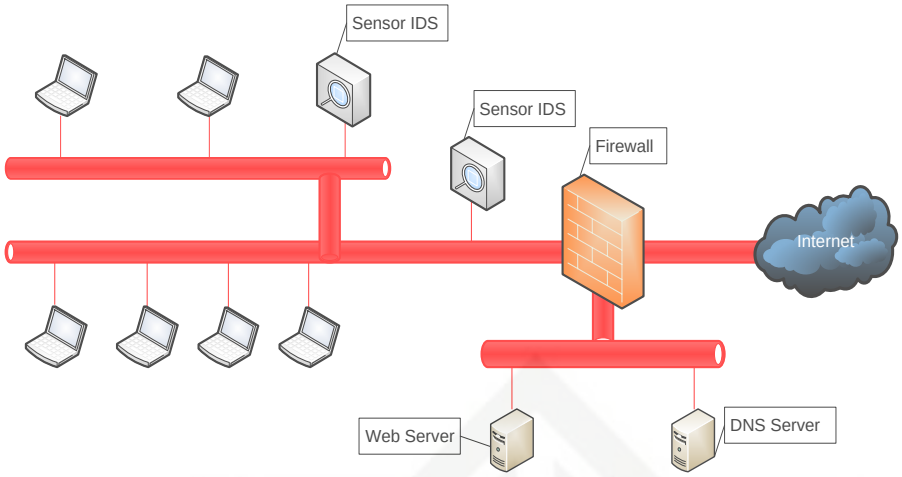
2.1.2 Jenis Intrusion Detection System (IDS)

Adapun jenis-jenis dari IDS yang bisa digunakan yaitu sebagai berikut:

a. *Network Based Inrusion Detection System* (NIDS)

Network based IDS adalah sebuah tipe IDS berbasis jaringan yang ditempatkan pada gateway yang berfungsi untuk melakukan pemeriksaan lalu lintas trafik jaringan yang mengalir melalui jaringan itu, baik yang menuju maupun berasal dari perangkat didalam jaringan. Sehingga semua perangkat yang berasal dari dalam maupun luar jaringan dilakukan *scanning* (Menaria, 2010).

Semua lalu lintas yang mengalir ke sebuah jaringan akan dianalisa paket header dan seluruh paket untuk mencari apakah ada percobaan sistem jaringan. NIDS umumnya terletak didalam segmen jaringan penting dimana server berada atau terdapat pada “pintu masuk” jaringan. Kelemahan NIDS adalah bahwa NIDS sedikit rumit diimplementasikan dalam sebuah jaringan yang menggunakan *switch ethernet*, meskipun beberapa *vendor switch ethernet* sekarang telah menerapkan fungsi IDS dalam *switch* buatannya untuk memonitor port atau koneksi (Hadi, dkk, 2012). Pada Gambar 2.1 merupakan contoh penerapan NIDS :



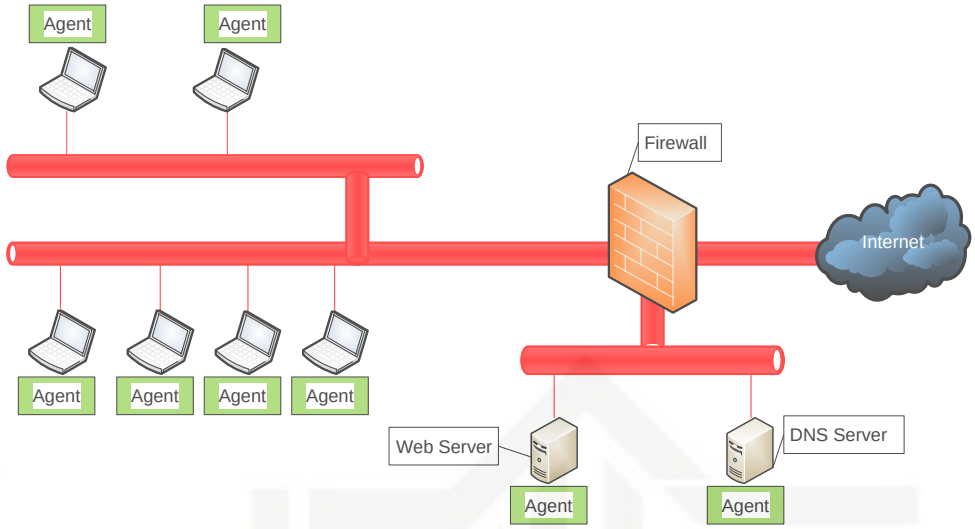
Gambar 2.1 Network Based IDS (Hadi, dkk, 2012)

b. *Host Based Intrusion Detection System* (HIDS)

Host Based IDS adalah sebuah system IDS yang berjalan pada *host* yang berdiri sendiri dalam sebuah jaringan. Tipe HIDS melakukan pengawasan terhadap trafik baik dalam maupun dari luar selanjutnya akan memberikan peringatan kepada *user* atau administrator jaringan ketika ada log yang mencurigakan (Menaria, 2010).

Kebanyakan produk IDS merupakan sistem yang bersifat pasif, mengingat tugasnya hanyalah mendeteksi intrusi yang terjadi dan memberikan peringatan kepada administrator jaringan bahwa mungkin ada serangan atau gangguan terhadap jaringan. Dalam hal ini, beberapa vendor sudah mengembangkan IDS yang bersifat aktif yang dapat melakukan beberapa tugas untuk melindungi *host* atau jaringan dari serangan ketika terdeteksi, seperti halnya menutup beberapa port atau memblokir beberapa alamt IP.

Produk seperti ini umumnya disebut sebagai *Intrusion Prevention System* (IPS). Beberapa produk IDS juga menggabungkan kemampuan yang dimiliki oleh HIDS dan NIDS yang kemudian disebut sebagai sistem hybrid (Hybrid Intrusion Detection System) (Hadi, dkk, 2012). Pada Gambar 2.2 merupakan contoh peneranan dari HIDS :



Gambar 2.2 Host Based IDS (Hadi, dkk, 2012)

Kelebihan dan Kekurangan IDS

Menurut Gondohanindijo, J., (2011), IDS memiliki beberapa kelebihan dan kelemahan diantaranya sebagai berikut:

a. Kelebihan IDS :

1. Memiliki akurasi keamanan yang baik dengan melakukan pendeteksian secara *realtime*.
2. Memiliki cakupan yang luas dalam mengenal proses attacking dan mampu mendeteksi segala sesuatu yang mencurigakan.
3. Dapat memberikan informasi tentang ancaman yang terjadi.
4. Memiliki tingkat forensik yang canggih dan mampu menghasilkan reporting yang baik.
5. Memiliki sensor yang dapat dipercaya untuk memastikan pendeteksian.

b. Kelemahan IDS :

1. Sering terjadi alarm atau gangguan yang bersifat palsu, yaitu paket yang datang terdeteksi sebagai intrusion karena tidak sesuai dengan rule yang dibuat. Setelah diteliti ternyata hanya paket data biasa dan tidak berbahaya.
2. False positives merupakan alert yang memberitahu adanya aktifitas yang berpotensi berupa serangan, tetapi masih ada kemungkinan bahwa aktifitas tersebut bukan sebuah serangan. Sehingga jika apabila jumlah alert banyak maka sulit untuk menyaring mana yang benar serangan atau bukan.
3. False negatives merupakan kondisi dimana IDS tidak dapat mendeteksi adanya serangan, karena tidak mengenal signature-nya. Sehingga IDS

tidak memberikan alert walaupun sebenarnya serangan terhadap system tersebut sedang berlangsung.

2.2 Intelligence Intrusion Detection Sistem (IIDS)

Intelligence Intrusion Detection System (IIDS) adalah suatu sistem deteksi intrusi baru yang dapat mendeteksi dan mengenali pola serangan baru dari serangan lama yang sudah ada, sistem secara otomatis membuat signature untuk serangan tersebut dan menambahkannya kedalam rule dimana secara sengaja memasang suatu kecerdasan buatan (Artificial Intelligence) ke dalam IDS (Anisyah, dkk, 2011).

Kecerdasan (Artificial Intelligence) buatan yang di pasang di IIDS ada 2 teknik atau metode dalam mendeteksi paket data yang berjalan pada jaringan yakni : *Anomaly-Based Detection* dan *Misused-Based Detection. Anomaly-Based Detection* adalah kemampuan mendeteksi bentuk intrusi jenis baru, ketika *anomaly* berbeda dari keadaan normal maka akan menrigger alarm, *Anomaly-Based Detection* terdapat beberapa algoritma dan model yang digunakan : *Data mining,* Algoritma Genetika dan *Artificial Intelligent*. *Anomaly-Based Detection* dapat dibagi dalam 3 kategori utama yaitu : *Statistical Based, Knowledge Based, Machine Learning Based. Misused-Based Detection* adalah mecocokkan data dengan deskripsi yang telah ditentukan dari perilaku (*behavior*) sebelumnya.

2.3 Algoritma K-Nearest Neighbor (K-NN)

Algoritma *k-nearest neighbor* (K-NN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut (Zainuddin, 2013). Rumus-rumus yang biasa digunakan sebagai ukuran jarak untuk data numerik ini antara lain:

a. *Euclidean Distance*

Ukuran ini sering digunakan dalam clustering karena sederhana.Ukuran ini memiliki masalah jika skala nilai atribut yang satu sangat besar dibandingkan nilai atribut lainnya.Oleh sebab itu, nilai-nilai atribut sering dinormalisasi sehingga berada dalam kisaran 0 dan 1.Untuk mendefinisikan jarak antara dua titik yaitu titik pada data *training* (x) dan titik pada data *testing* (y) maka digunakan rumus *Euclidean*, seperti yang ditunjukkan pada berikut:

$d(x,y) = \sqrt{\sum_{i=1}^{n}(xi - yi)^2}$ …….. (2.1)

dengan :

d : jarak antara titik pada data *training* x dan titik data *testing* y yang akan diklasifikasi, dimana x=x1,x2,…,xi dan y=y1,y2,…,yi\

x : data uji

y : data latih

I : merepresentasikan nilai atribut

n : merupakan dimensi atribut.

b. *City Block Distance*

Jika tiap item digambarkan sebagai sebuah titik dalam grid, ukuran jarak ini merupakan banyak sisi yang harus dilewati suatu titik untuk mencapai titik yang lain seperti halnya dalam sebuah peta jalan.

c. *Manhattan Distance*

Manhattan Distance merupakan salah satu pengukuran yang paling banyak digunakan meliputi penggantian perbedaan kuadrat dengan menjumlahkan perbedaan absolute dari variable-variable. Fungsi ini hanya akan menjumlahkan selisih nilai x dan y dari dua buah titik.

d. *Minkwoski Metric*

Ukuran ini merupakan bentuk umum dari *Euclidean Distance* dan *Manhattan Distance*.*Euclidean Distance* adalah kasus dimana nilai p=2 sedangkan *Manhattan Distance* merupakan bentuk *Minkwoski* dengan p=1. Dengan demikian, lebih banyak nilai numerik yang dapat ditempatkan pada jarak terjauh di antara 2 vektor.Seperti pada *Euclidean Distance* dan juga *Manhattan Distance*, ukuran ini memiliki masalah jika salah satu atribut dalam vektor memiliki rentang yang lebih besar dibandingkan atribut-atribut lainnya.

d. *Cosine*

Ukuran ini bagus digunakan pada data dengan tingkat kemiripan tinggi walaupun sering pula digunakan bersama pendekatan lain untuk membatasi dimensi dari permasalahan. Dalam mendefenisikan ukuran jarak antar k yang digunakan beberapa algoritma untuk menentukan k mana yang terdekat.

K-NN adalah algoritma untuk mengklasifikasi objek baru berdasarkan atribut dan *training samples* (data latih). Dimana hasil dari sampel uji yang baru

diklasifikasikan berdasarkan mayoritas dari kategori pada *K-NN*. Algoritma *K-NN* menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru (Krisandi, N. dkk, 2013). Data latih akan dibangun dengan memperhatikan keseimbangan dokumen satu sama lain. Adapun algortima *K-NN* dapat dijelaskan dengan keterangan berikut :

1. Menentukan parameter nilai *k* = jumlah tetanggaan terdekat.

   $K = \sqrt{n}$.....(2.2)
2. Hitung jarak antara data uji dengan semua data latih menggunakan rumus *Euclidean Distance* 2.1.
3. Urutkan jarak yang terbentuk dari keci ke besar dan tentukan jarak terdekat sampai urutan ke-k.
4. Pasangkan kategori sesuai atau kelas yang bersesuaian.
5. Cari jumlah terbanyak dari tetanggaan terdekat dengan persamaan 2.2. Kemudian tetapkan kategori.

   Jarak yang digunakan dalam penelitian ini adalah *Euclidean Distance.*

Algoritma *K-NN*(Krisandi, N. dkk, 2013) adalah algoritma yang menentukan nilai jarak pada pengujian data *testing* dengan data *training* berdasarkan nilai terkecil dari nilai ketetanggaan terdekat.

2.1.3 Confusion Matrix

Confusion matrix adalah alat visualisasi yang digunakan pada *supervised learning.* Tiap kolom pada matrix adalah contoh kelas prediksi sedangkan tiap baris mewakili kejadian dikelas yang sebenarnya. *Confusion matrix* berisi informasi *actual* dan prediksi pada sistem klasifikasi (Amir, 2015).

Pengujian *confusion matrix* menggunakan tiga kategori yaitu *precision, recall* dan *accuracy*. Berikut perhitungan mencari nilai *precision, recall* dan *accuracy* yang dapat kita lihat pada Tabel 2.1:

Tabel 2.1 Contoh *confusion matrix*

		Predicted Calss	
		Serangan	Tidak serangan
Actual Class	Serangan	a	b
	Tidak serangan	c	d

Precision : $\frac{d}{b+d}$

Recall : $\frac{d}{c+d}$

Accuracy: $\frac{a+d}{a+b+c+d}$ .....(2.3)

Keterangan:

1) a bernilai benar, karena data serangan menghasilkan data serangan.
2) b bernilai salah, karena data serangan menghasilkan data tidak serangan.
3) c bernilai salah, karena data tidak serangan menghasilkan data serangan.
4) d bernilai benar, karena data tidak serangan menghasilkan data tidak serangan.
5) *Precision* adalah proporsi kasus positif yang diidentifikasi dengan benar.
6) *Recall* adalah proporsi kasus positif yang diidentifikasi dengan benar.

Accuracy adalah perbandingan kasus yang diidentifikasi benar dengan jumlah semua kasus.

2.4 Penelitian Terkait

Penelitian sebelumnya IIDS dengan menggunakan metode *KNN* cukup banyak, dengan menyelesaikan kasus-kasus yang berbeda dapat dilihat pada Tabel 2.2 berikut ini :

Tabel 2.2 Penelitian Terkait

Penulis	Judul penelitian	Kesimpulan
Wahyu Nugroho 2014	Rancang Bangun Aplikasi *Intrusion Detection System* Dengan Menggunakan Metode *Fuzzy*	Dari uji coba yang dilakukan penelitian ini mampu mengindetifikasi dan mengklasifikasikan serangan paket normal, paket UDP flooding, paket TCP flooding, paket ICMP flooding. Namun pada penelitian ini belum mampu mengindetifikasi dan mengklasifikasikan paket serangan Syn-Ack dan HTTP.

Penulis	Judul penelitian	Kesimpulan
Ririn Dwi Jayanti, Noeryanti 2014	Aplikasi Metode K-Nearest Neighbor Dan Analisa Diskriminan Untuk Analisis Resiko Kredit Pada Koperasi Simpan Pinjam Di Kopinkra Sumber Rejeki	Kesimpulan yang diperoleh dari hasil penelitian variabel jangka waktu berkorelasi negatif terhadap kelancaran dan ketetapan prediksi resiko kredit yang dianalisis dengan metode K-NN sebesar 84,33% pada nilai k adalah 7 sedangkan ketetapan prediksiresiko kredit yang dianalisis dengan analisis Diskriminan adalah sebesar 76,5%.
Novi Anisyah 2011	Aplikasi mobile untuk *K-nearest neoghbor (K-NN)* pada intrusion detection sistem berbasis snort	Berdasarkan hasil eksprimen, penelitian ini mendapatkan hasil bahwa metode KNN sudah baik untuk mendeteksi serangan jika dibandingkan dengan fuzzy maka semakin besar data base dan semakin kecil K makan persen bedanya semakin kecil.
Raymundus, dkk	Sistem pendukung keputusan untuk menentukan status prestasi siswa menggunakan metode *K-Nearest Neighbor*	Hasil penelitian dalam menetukan sistem pendukung keputusan untuk menentukan prestasi siswa khususnya pada uji coba ditabel 1 yaitu kelas 1. Misalnya k=5 sehingga status prestasi siswa, yaitu Billy adalah rendah.
Akhmad Alimudin 2013	Sistem Deteksi Intrusi Pada Jaringan Dengan Menggunakan Metode *K-Nearest Neighbor* dan Teori *Dempster Shafer*.	Hasil penelitian menunjukkan bahwa, penggunaan metode K-NNdengan teori dempster shafer mendapatkan tingkat akurasi yang lebih tinggi dibandingkan dengan metode *Distance Weighted*dan *voting*, namu tingkat akurasi DSK-NN juga dipengaruhi oleh banyaknya K dan jarak yang dihasilkan.