Evaluasi Sistem Pendeteksi Intrusi Berbasis Anomali dengan N-Gram dan

Incremental Learning

I Made Agus Adi Wirawan, Royyana Muslim Ijtihadie, dan Baskoro Adi Pratomo  
Jurusan Teknik Informatika, Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember (ITS)  
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia  
*e-mail*: {roy, baskoro}@if.its.ac.id

*Abstrak*—Perkembangan data digital cukuplah pesat, mulai dari ukuran hingga jenis berkas. Untuk mengimbangi hal tersebut maka perlu ada-nya algoritma pemampatan guna mengurangi ukuran penggunaan media penyimpanan. Salah satu algoritma yang cukup lama ada adalah *Adaptive Huffman*, dan serta mengkombinasikannya dengan *LZW*, maka hasil pemampatan yang didapat akan lebih hemat. Pada paper ini kami mencoba membuktikan hal tersebut dan juga dengan menambahkan sebuah algoritma pengolahan awal untuk berkas citra guna bersifat *loss*y. Algoritma pengolahan awal yang akan kami jabarkan menggunakan metode mean, median dan modus untuk menghitung dan menentukan nilai *LSB* baru secara adaptif. Dari hasil percobaan didapatkan rerata nilai *PSNR* sebesar 36,26 dB dan rerata rasio pemampatan sebesar 44*,*77 % pada metode median.

*Kata Kunci*—*N-Gram*, Mahalanobis Distance, Incremental Learning.

# PENDAHULUAN

S

EMAKIN pesatnya perkembangan teknologi informasi memudahkan orang-orang untuk saling tukar menukar data baik melalui internet maupun intranet. Tentunya dengan mudahnya berbagi data itulah sangat memungkinkan terjadinya serangan terhadap data tersebut terutama melalui jaringan komputer. Sistem pendeteksi intrusi atau yang pada umumnya disebut IDS (*Intrusion Detection System*) merupakan senjata utama untuk mengamankan suatu jaringan dimana sistem ini nantinya bertugas untuk mengidentifikasi dan mencatat apakah suatu paket data tersebut merupakan bentuk serangan atau paket data bisa.

Saat ini telah banyak dikembangkan aplikasi IDS (*Intrusion Detection System*), namun sebagian besar yang dikembangkan berbasis *signature* atau menggunakan *rule*, dan sebagian kecil menggunakan *anomaly*. *Anomaly* pada dasarnya adalah mencari data yang menyimpang dari sekumpulan data normal. IDS yang berbasis pada *anomaly* bersifat lebih fleksibel, karena dapat mengenali pola serangan baru tanpa harus memperbaharui basis data pola serangan. IDS yang berbasis pada anomali memiliki sebuah kecerdasan buatan yang mampu mendeteksi dan mengenali sebuah serangan. IDS yang berbasis anomali menggabungkan metode analisis dan statistik untuk mengenali penyimpangan tersebut. Kelemahan dari metode ini adalah kemungkinan salah identifikasi pada data yang diolah.

Sistem kerja intrusi ini pada dasarnya dikirimkan lewat jaringan dengan paket-paket data yang sama dengan paket data normal. Dengan banyaknya paket data yang masuk kedalam sebuah host, tentunya host ini harus bisa mengenali paket data, apakah paket data tersbut terdapat paket data yang berupa intrusi atau tidak. Hal tersebut dapat dikenali dengan cara mengelompokkan data berdasarkan beberapa hal yang membedakan antara paket data normal dengan paket data yang berupa intrusi.

Maka untuk membedakan hal tersebut diperlukan sebuah sistem deteksi intrusi dimana nantinya sistem deteksi intrusi tersebut menggunakan gabungan metode analisis dan statistik yang berfungsi mengenali perbedaan paket data normal maupun paket data berupa intrusi. Selain itu, sistem deteksi intrusi yang dapat mempelajari paket data normal yang baru sebagai data *training*.

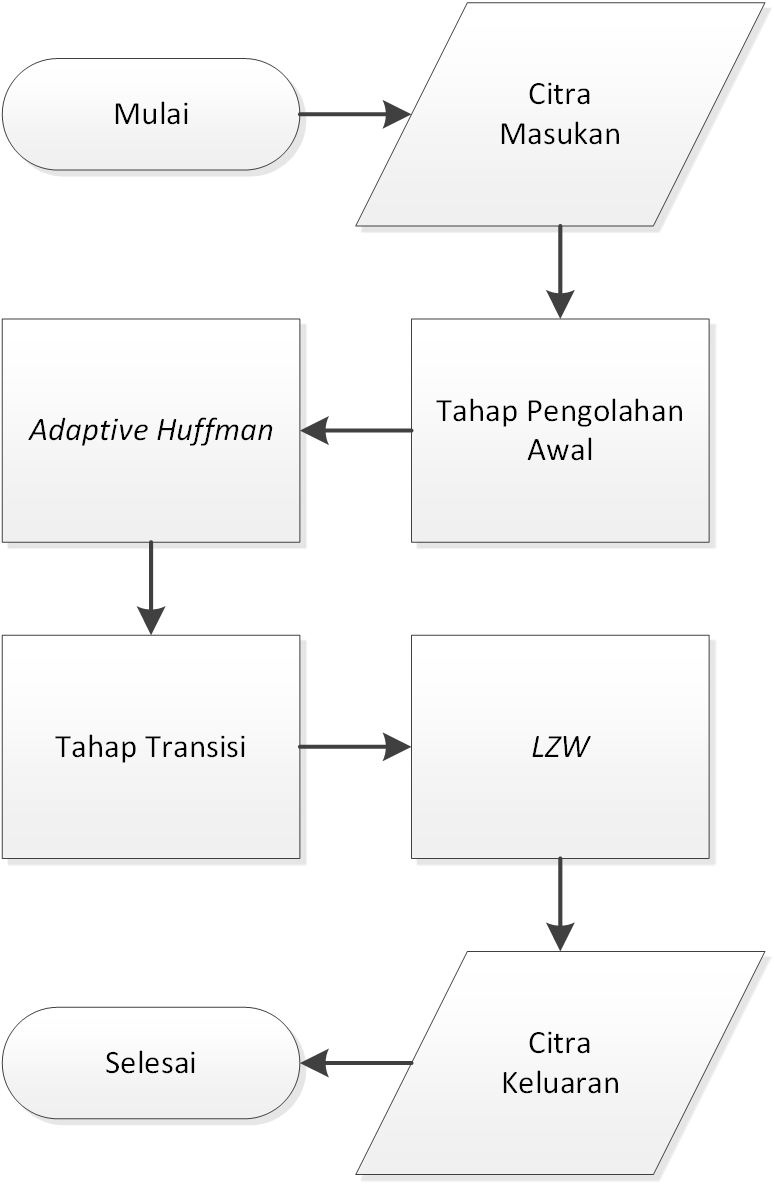
Untuk dapat menghitung jarak mahalanobis dari paket data, diperlukan metode yang dapat merubah informasi paket data menjadi nilai yang dapat dihitung. Metode n-gram dapat digunakan untuk mebuat model yang sederhana dan cepat untuk dihitung khususnya menghitung distribusi karakter pada suatu paket data. metode n-gram merupakan metode yang paling efesien dan efektif dalam membuat model dari suatu paket data.

Metode mahalanobis *distance* berguna untuk membedakan paket-paket data berdasarkan anomali yang terjadi. Untuk dapat mempelajari paket data normal yang baru menggunakan metode *incremental learning*, dimana metode ini nantinya memperbaharui rata-rata dan standar deviasi dari model paket data yang ada pada data *training*.

Artikel ini menggunakan artikel [4] sebagai acuan, dan mengimplementasikan algoritma pemampatan citra secara *lossy* dan berbasis *dictionary* dengan menggunakan *Adaptive Huffman* dan *LZW*, serta dengan mengubah *Last Significant Bit* (*LSB*) agar berkas citra dapat dikemas menjadi sebuah berkas baru yang memiliki ukuran lebih hemat dibandingkan berkas citra awal. Susunan artikel ini adalah sebagai berikut: 1 Pendahuluan, 2 Desain Algoritma, 3 Pengujian, dan 4 Kesimpulan.

# DESAIN ALGORITMA

Alur algoritma pemampatan ini terdiri atas empat tahap sesuai dengan diagram alur yang tertera pada Gambar 1.



Gambar. 1. Diagram Alur Algoritma Pemamatan

## Pengolahan Awal

Menggunakan metode mean, median dan modus untuk menghitung nilai distribusi. Nilai distribusi digunakan untuk menentukan nilai *LSB* pengganti. Dengan ketentuan sesuai dengan persamaan 1 dan 2.

(1)

(2)

Dimana *F(dist)* adalah nilai distribusi yang didapatkan setelah melakukan penghitungan dengan metode mean, median maupun modus. Tidak dengan ketiganya sekaligus.

## Adaptive Huffman

*Adaptve Huffman* yang digunakan mengacu pada artikel [5]. Dengan menggunakan jumlah simbol = 28 atau 256 karakter. Pada tahap ini berkas citra akan dimampatkan secara linear atau dengan mengubah bentuk dari dua dimensi menjadi satu dimensi.

## Tahap Transisi

Pada tahap ini, aliran dantum hasil pemampatan dengan *Adaptive Huffman* akan dipotong-potong sesuai dengan nilai modus kedalaman Pohon *Huffman* yang terbentuk. Dengan melakukan pemotongan ini maka hasil pemampatan dengan *Adaptive Huffman* dapat dimampatkan kembali secara maksimal oleh *LZW*. Pemotongan ini sendiri memiliki batasan sesuai dengan persamaan 3.

Gambar. 2. Diagram Perbandingan Rasio Pemampatan dengan Menggunakan Metode Mean, Median dan Modus

(3)

Dimana *s* adalah ukuran pemotongan sepanjang *s* bit, dan *F(s)* adalah ukuran terpilih.

## LZW

*LZW* yang digunakan mengacu pada artikel [6], dan struktur data yang digunakan mengacu pada artikel [7]. Dengan menggunakan ukuran kamus data sesuai persamaan 4.

Gambar. 3. Diagram Perbandingan Nilai *PSNR* dengan Menggunakan Metode Mean, Median dan Modus

(4)

Dimana *l* adalah ukuran kamus data sepanjang *l* bit, dan *F(l)* adalah ukuran terpilih.

# PENGUJIAN

## Skenario Pengujian

Pengujian yang dilakukan dibagi kedalam tiga bagian, yakni:

1. Pengujian mean, median dan modus;
2. Pengujian alur transisi; dan
3. Pengujian penentuan nilai *LSB*.

Parameter pengujian yang digunakan adalah rasio pemampatan dan *PSNR*. Dimana rasio pemampatan yang digunakan sesuai dengan persamaan 5.

(5)

Dengan ketentuan bahwa semakin kecil rasio semakin baik dan semakin besar nilai *PSNR* semakin baik pula. Sebuah citra *lossy* dikatakan baik jika nilai *PSNR* yang dimiliki lebih besar dari 40 *dB* dan dikatakan cukup jika dibawah 30 *dB* [8].

## Citra Pengujian

Citra yang digunakan menggunakan tipe berkas *PGM P5*. Citra-citra tersebut dapat diunduh atau diakses pada laman http://www.ece.rice.edu/~wakin/images/ dan http://www.csd .uwo.ca/courses/CS4487a/image\_samples.html.

Citra-citra tersebut antara lain:

* lena512.pgm
* bell.pgm
* class57.pgm
* voyager2.pgm

## Hasil Pengujian

### Pengujian Metode Mean, Median dan Modus

Hasil pengujian menunjukan *trade-off* muncul pada metode median, dimana rerata rasio pemampatan yang diperoleh adalah 44,77 % dan rerata *PSNR* sebesar 36,26 dB. Grafik hasil pengujian dapat dilihat pada Gambar 2 dan 3.

### Pengujian Alur Transisi

Hasil pengujian menunjukkan bahwa dengan menambahkan alur transisi *LZW* secara umum membuat hasil pemampatan menjadi lebih mangkus, dimana rasio pemampatan menjadi lebih padat dari awalnya memiliki rerata 44,77 % menjadi 38,99 %. Grafik hasil pengujian dapat dilihat pada Gambar 4. Hasil juga menunjukkan bila penggunakan tahap transisi atau pemotongan aliran dantum adalah langkah tepat.

Gambar. 4. Diagram Perbandingan Rasio Pemampatan Dengan dan Tanpa *LZW* serta dengan Tahap transisi

### Pengujian Penentuan Nilai LSB

Hasil pengujian menemukan nilai *trade-off* dari nilai *PSNR* dan ukuran rasio dengan berlandaskan pada nilai distribusi diantara 5 % hingga 15 % dan dipilih yang paling kecil. Batasan tersebut didapat dengan menguji setiap citra dengan nilai panjang *LSB* mulai dari 1 hingga 8 bit. Grafik hasil pengujian dapat dilihat pada Gambar 5 dan 6.

Gambar. 6. Diagram Perbandingan Nilai *PSNR* untuk Setiap *LSB*

Gambar. 5. Diagram Perbandingan Rasio Pemampatan untuk Setiap *LSB*

# KESIMPULAN

Citra keabuan dapat dimampatkan dengan menggunakan penggabungan *Adaptive Huffman* dan *LZW*, serta akan bersifat *lossy* dengan melalui tahap pengolahan awal. Dengan adanya tahap transisi, pemampatan oleh *LZW* dapat dimaksimalkan. Selain itu dengan membatasi ukuran kamus data dan melakukan *flushing* kamus data, maka ledakan kamus data pada *LZW* dapat dicegah.

DAFTAR PUSTAKA

|  |  |
| --- | --- |
| [1] | K. Sayood, Introduction to Data Compression Third Edition, San Francisco: Elsevier, 2006. |
| [2] | V. Crnojevic, V. Senk, Z. Trpovski, "Lossy Lempel-Ziv algorithm for image compression," *Telecommunications in Modern Satellite, Cable and Broadcasting Service,* vol. 2, 2003. |
| [3] | Gabriela Dudek, Przemysław Borys, Zbigniew J. Grzywna, "Lossy dictionary-based image compression method," *Image and Vision Computing,* vol. 25, no. 6, pp. 883-889, 2007. |
| [4] | Dalvir Kaur and Kamaljit Kaur, "Huffman Based LZW Lossless Image Compression Using Retinex Algorithm," *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 2, no. 8, pp. 3145-3151, 2013. |
| [5] | J. S. Vitter, "Design and Analysis of Dynamic Huffman Codes," *Journal of the ACM,* vol. 34, no. 4, pp. 825-845, 1987. |
| [6] | T. A. Welch, "A Technique for High-Performance Data Compression," *IEEE Computer,* vol. 17, no. 6, pp. 8-19, 1984. |
| [7] | E. Fredkin, "Trie memory," *Communications of the ACM,* 1960. |
| [8] | Cheddad, A., Condell, J., Curran, K., Kevitt, P.Mc, "Digital Image Steganography : Survey and Analysis of Current Methods," *Signal Processing,* vol. 90, no. 3, pp. 727-752, 2010. |