JITE (Journal of Informatics and Telecommunication Engineering)

Available online <http://ojs.uma.ac.id/index.php/jite> DOI : 10.31289/jite.v4i1.3873

Received: 14 Juni 2020

Accepted: 07 Juli 2020

Published: 20 Juli 2020

Detection of Attacks on Apache2 Web Server Using Genetic Algorithm Based On Jaro Winkler Algorithm

M Rizqi Maulana^{1)*}

1) Prodi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sukabumi, Indonesia

*Corresponding Email: rizqimaaulana69@gmail.com

Abstrak

Web server merupakan perangkat lunak yang memberikan layanan data berupa request HTTP (*Hypertext Transfer Protocol*) dan respon berupa dokumen HTML (*Hypertext Markup Language*) dengan tujuan untuk mengelola data berupa berkas teks, gambar, video, dan file. Namun dalam mengelola data dalam jumlah yang besar, dibutuhkan monitoring keamanan yang baik agar data yang disimpan di dalam web server tidak mudah diretas. Untuk melindungi web server dari peretas perlu adanya aplikasi untuk mendeteksi aktifitas yang dianggap mencurigakan atau kemungkinan aktifitas peretasan. Dengan memanfaatkan log dari web server yang diolah menggunakan algoritma Jaro Winkler untuk melihat upaya peretasan yang menghasilkan matrik dan laporan aktifitas peretasan kepada admin. Dengan demikian admin web server dapat melihat aktifitas mencurigakan pada web server secara langsung.

Kata Kunci: Web Server, Algoritma Jaro Winkler.

Abstract

Web server is software that provides data services in the form of HTTP (*Hypertext Transfer Protocol*) requests and responses in the form of HTML documents (*Hypertext Markup Language*) with the aim of managing data in the form of text files, images, videos and files. But in managing large amounts of data, good security monitoring is needed so that the data stored on the web server is not easily hacked. To protect the web server from hackers need an application to detect activities that are considered suspicious or possible hacking activities. By utilizing logs from a web server that is processed using the Jaro Winkler algorithm to see hacking attempts that produce a matrix and hacking activity reports to the admin. Thus the web server admin can see suspicious activity on the web server directly.

Keywords: Web Server, Jaro Winkler Algorithm.

How to Cite: Maulana, M.R. (2020). Detection of Attacks on Apache2 Web Server Using Genetic Algorithm Based On Jaro Winkler Algorithm. *JITE (Journal Of Informatics And Telecommunication Engineering)*. 4 (1): 185-192

I. PENDAHULUAN

Teknologi informasi dan komunikasi pada setiap perusahaan bisnis digunakan di semua aspek baik profit maupun nonprofit (Fahmi, Fachrudin & Silvia, 2019). Hal itu mengakibatkan penggunaan aplikasi berbasis teknologi informasi di setiap perusahaan semakin meningkat. Sebagai contoh saat ini banyak aplikasi bisnis seperti e-commerce, mobile banking dan lain sebagainya (Rahardjo 2005). Penggunaan aplikasi berbasis teknologi informasi memicu meningkatnya kejahatan komputer. Server yang digunakan untuk menyimpan data perusahaan harus diimbangi dengan monitoring yang baik oleh administrator handal. Dengan banyak nya server yang harus dilindungi, administrator sering kesulitan untuk memonitoring setiap server. Akibat dari hal ini server sering kali diretas oleh seorang cracker (Rahardjo 2005).

Menurut Raharjo, (2015) Serangan terhadap keamanan sistem informasi dapat dilihat dari beberapa aspek berikut :

1. *Interruption*: serangan ditujukan untuk merusak perangkat sistem. Contoh serangan adalah *Danial of Service Attack*.
2. *Interception*: serangan berhasil mengakses data atau informasi.

Contohnya adalah penyadapan (*Wiretapping*)

3. *Modification*: serangan berhasil mengakses serta mengubah data dan informasi.
4. *Fabrication*: serangan ditujukan untuk menyisipkan objek palsu ke dalam sistem.

Similarity metric adalah matriks berbasis teks dengan hasil berupa nilai kesamaan atau ketidaksamaan dari dua buah string (Kurniawati, Puspitodjati, and Rahman 2010). Dengan melihat nilai Similarity metric antara log server normal dan kumpulan data log server annormal semakin tinggi nilai Similarity metric maka semakin tinggi tingkat kemungkinan peretasan.

Jaro-Winkler distance merupakan algoritma untuk mengetahui tingkat kesamaan dua string. Semakin tinggi nilai Jaro-Winkler distance antarta dua string, maka tingkat kesamaan antara dua string semakin tinggi. Skor normal dari Jaro-Winkler distance adalah 0 menandakan tidak sama dan 1 menandakan kesamaan (Kurniawati, Puspitodjati, and Rahman 2010).

Berdasarkan identifikasi permasalahan tersebut maka rumusan masalahnya adalah bagaimana mengolah data log yang berisi data aktifitas dari web server agar menjadi informasi untuk

mendeteksi adanya usaha peretasan pada sistem.

Algoritma genetika merupakan suatu teknik pencarian atau teknik optimasi dengan konsep evolusi struktur genetika pada makhluk hidup (Arkeman, Seminar, and Gundawan 2012).

Nilai fitness merupakan tahapan dalam algoritma genetika dengan menentukan nilai masing masing individu dengan kriteria sesuai permasalahan penelitian (Mahmudy 2014).

Crossover adalah mekanisme dalam algoritma genetika dengan menggabungkan dua kromosom sehingga menghasilkan kromosom baru (Mahmudy 2013).

Proses mutasi adalah menyiptakan individu baru dengan modifikasi satu atau lebih gen dalam individu yang sama (Zukhri 2004). Algoritma genetika diterapkan agar proses pencarian nilai Jaro-Winkler distance jadi lebih optimal.

Pada penelitian kali ini algoritma jaro-winkler distance dan algoritma genetika diterapkan untuk mencari nilai kesamaan dan ketidaksamaan pada log web server. Batasan pada penelitian kali ini dapat dilihat sebagai berikut :

1. *Web server* yang digunakan adalah *Apache2.4*

2. Bahasa pemrograman yang digunakan yaitu *python* versi 3.8 dengan dukungan database *mysql*.

Tujuan dari penelitian kali ini yaitu untuk membuat aplikasi yang dapat melihat aktifitas server dan mendeteksi aktifitas yang dianggap berpotensi melakukan peretasan terhadap sistem agar administrator web server dapat memonitoring aktifitas server dengan lebih mudah.

II. METODE PENELITIAN

A. *Analisis*

Optimasi algoritma *Jaro-Winkler* menggunakan algoritma genetika dapat digunakan untuk mengembangkan *log web server*. Berikut alur penelitian kali ini :

Gambar 1 Diagram alur

Aturan yang dibuat digunakan untuk mengetahui koneksi normal dan koneksi tidak normal. Aturan aturan bisa didefinisikan sebagai berikut :

$$\text{If} \{ \text{condition} \} \text{ then } \{ \text{act} \} \quad (1)$$

Dari aturan diatas, kondisi pada aturan tersebut mengacu pada hasil nilai kesamaan dan ketidaksamaan antara log server saat ini dan basis pengetahuan log server tidak normal yang diperoleh melalui algoritma jaro-winkler. Aturan yang digunakan pada algoritma jaro-winkler adalah sebagai berikut,

$$d_j = \frac{1}{3} \times \left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m} \right) \quad (2)$$

Dimana :

m = jumlah karakter yang sama persis

$|s_1|$ = panjang string 1

$|s_2|$ = panjang string 2

t = jumlah transposisi

Hasil dari dua string yang diolah akan menghasilkan nilai 0 hingga 1, 0 untuk nilai tidak sama dan 1 untuk nilai sama persis. Data log web server yang diolah berupa IP Address, status, userAgent, query, request, byte_in, byte_out. Dari hasil nilai jaro-winkler dan data log berikut maka aturan selanjutnya dapat didefinisikan sebagai berikut,

*If { value of jaro winkler status = 1,
value of jaro winkler userAgent = 1,
value of jaro winkler query = 1,
value of jaro winkler request = 1,
(3)
value of jaro winkler bytes_in = 1,
value of jaro winkler bytes_out = 1 }
then { block ip address }*

aturan diatas dapat diartikan apabila nilai jaro-winkler dari jaro-winkler status sama dengan 1, nilai dari jaro-winkler userAgent sama dengan 1, nilai dari jaro winkler query sama dengan 1, nilai dari jaro winkler request sama dengan 1, nilai jaro winkler bytes_in sama dengan 1, nilai jaro winkler bytes_out sama dengan 1 maka block ip address.

B. Represesntasi Populasi

Untuk melihat log normal dan log tidak normal dari web server perlu menentukan populasi dengan melihat data log secara keseluruhan dengan koneksi tertentu dan diambil secara acak, dari hasil tersebut peneliti memilih beberapa atribut sebagai berikut :

Tabel 1 Atribut log server

Atribut
Status
UserAgent
Query
Request
Bytes_in
Bytes_out

C. Menentukan Kromosom dan Mutasi

Dari data populasi tersebut diambil satu koneksi yang diambil secara urut dan dijadikan individu. Individu dirubah menjadi bentuk kromosom sebagai berikut:

[0, 1, 0, 1, 1, 1]

Terdapat tujuh gen pada kromosom gen pertama pada kromosom mewakili atribut status, gen ke dua pada kromosom menentukan nilai UserAgent berikut seterusnya sesuai dengan atribut yang terdapat pada populasi.

Nilai 1 dan 0 di dalam kromosom didapat dari hasil nilai jaro-winkler setiap atribut dengan aturan sebagai berikut,

```
if { condition } then
    { return: 1 }
else
    { return: 0 }
```

(4)

Kondisi pada aturan diatas menjelaskan bahwa ketika kondisi terpenuhi maka nilai kembali adalah 1, apabila tidak terpenuhi maka nilai kembali adalah 0. Kondisi untuk setiap aturan dapat berbeda beda berikut aturan untuk menentukan kondisi atribut status,

```
if { status > 0.7 } then
    { return: 1 } else
    { return: 0 }
```

(5)

Aturan untuk menentukan kondisi dari atribut *userAgent*, *query* dan *request* adalah sebagai berikut:

```
if { userAgent, query, request > 0.4 }
then
```

(6)

```
{ return: 1 }
else
    { return: 0 }
```

Aturan untuk menentukan kondisi dari atribut *bytes_in* adalah sebagai berikut:

```
if { bytes_in > 38891 } then
(7)
    { return: 1 } else
    { return: 0 }
```

Aturan untuk menentukan kondisi dari atribut *bytes_out* adalah sebagai berikut:

```
if { bytes_in > 7320 } then
(8)
    { return: 1 } else
    { return: 0 }
```

D. Menghitung Nilai Fitness

Menghitung nilai fitness merupakan tahapan yang sangat penting. Nilai fitness dicari dengan menjumlahkan seluruh nilai gen pada setiap individu kromosom. Hasil dari nilai fitness akan dievaluasi untuk melihat koneksi normal dan koneksi tidak normal. Nilai fitness dapat diperoleh dengan menjumlahkan seluruh nilai gen yang ada disetiap kromosom. Selengkapnya cara menghitung nilai fitness bisa dilihat pada contoh berikut :

Kromosom = [1, 0, 1, 0, 1, 1, 1]

$$Fitness = 1 + 0 + 1 + 0 + 1 + 1 + 1$$

$$Fitness = 5$$

Dari contoh diatas terlihat bahwa nilai fitness 5 didapat dari jumlah seluruh dengan nilai 1, karena gen dengan nilai 1 ada 5 maka nilai fitness dari kromosom tersebut adalah 5.

III. HASIL DAN PEMBAHASAN

A. Sample Data

Sample data yang digunakan pada penelitian kali ini berjumlah 100 log web server. Berikut lampiran contoh koneksi - log web server :

Tabel 2 contoh koneksi

Atribut	Value
Koneksi 1	
Status	200
UserAgent	sqlmap/1.4.5.24#dev (http://sqlmap.org)
Query	AND 3464=9707 AND ('ZSAu'='ZSAu
Request	AND 3464=9707 AND ('ZSAu'='ZSAu
Bytes_in	15
Bytes_out	0
Koneksi 2	
Status	500
UserAgent	-
Query	AND 3464=9707 AND ('ZSAu'='ZSAu
Request	/admin/pocustomer/ view_pocustomer_edit/Ng==
Bytes_in	15
Bytes_out	0
Koneksi 3	
Status	500
UserAgent	Mozilla/5.0 (X11; Linux x86_64)
Query	AND 3464=9707 AND ('ZSAu'='ZSAu
Request	-
Bytes_in	15
Bytes_out	0
Koneksi 4	
Status	500
UserAgent	sqlmap/1.4.5.24#dev (http://sqlmap.org)
Query	AND 3464=9707 AND ('ZSAu'='ZSAu
Request	AND 3464=9707 AND ('ZSAu'='ZSAu

Bytes_in	15
Bytes_out	0

Tabel 3 contoh koneksi tidak normal

Atribut	Value
Status	500, 400, 404
UserAgent	Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/81.0.4044.129 Safari/537.36
Query	AND 3464=9707 AND 3464=9707 AND ('ZSAu'='ZSAu AND (('ZSAu'='ZSAu AND 8554= (SELECT UPPER (XMLType(CHR(60) CHR(58) CHR(113) CHR(106) CHR(107) CHR(112) CHR(113) (SELECT (CASE WHEN (8554=8554) THEN 1 ELSE 0 END) FROM DUAL) CHR(113) CHR(106) CHR(122) CHR(98) CHR(113) CHR(62))) FROM DUAL) AND (('wmKE'='wmKE
Request	AND 3464=9707 AND 3464=9707 AND ('ZSAu'='ZSAu AND (('ZSAu'='ZSAu AND 8554= (SELECT UPPER (XMLType(CHR(60) CHR(58) CHR(113) CHR(106) CHR(107) CHR(112) CHR(113) (SELECT (CASE WHEN (8554=8554) THEN 1 ELSE 0 END) FROM DUAL) CHR(113) CHR(106) CHR(122) CHR(98) CHR(113) CHR(62))) FROM DUAL) AND (('wmKE'='wmKE
Bytes_in	38891
Bytes_out	7320

B. Hasil Penentuan Kromosom dan Mutasi

Menentukan nilai gen pada kromosom diawali dengan menentukan nilai *jaro-winkler distance* untuk setiap atribut yang ada pada log web server. Nilai dari *jaro-winkler distance* berupa angka desimal dengan rentan dari angka 0 hingga angka 1. Nilai yang mendekati angka 0 maka dapat dinyatakan bahwa string yang

dibandingkan tidak sama sedangkan nilai jaro-winkler yang mendekati 1 dapat dinyatakan sama. Berikut hasil nilai jaro-winkler distance :

```

/home/rizqimaulana/PycharmProjects/guessPassword/venv/bin/python /home/r
[1.0, 0.44351, '1', 0.646023, 0.73639]
[0.777778, 0.0, '1', 0.252218, 0.73639]
[0.777778, 0.962963, '1', 0.0, 0.73639]
[0.777778, 0.44351, '1', 0.646023, 0.73639]
[0.777778, 0.44351, '1', 0.58679, 0.73639]

Process finished with exit code 0

```

Gambar 2 Hasil nilai jaro-winkler distance

Dari gambar berikut terdapat nilai dari hasil perbandingan antara log normal dan log tidak normal. Nilai dari jaro-winkler distance dirubah menjadi 0 dan 1, nilai tersebut merupakan gen dari setiap kromosom. Berikut hasil dari kromosom :

```

['0', '1', '1', '1', '1']
['1', '1', '1', '0', '1']
['1', '0', '1', '0', '1']
['1', '1', '1', '1', '1']
['1', '1', '1', '1', '1']

```

Gambar 3 Bentuk kromosom

C. Hasil Nilai Fitness

Hasil dari nilai *fitness* akan menentukan apakah koneksi tersebut termasuk normal atau tidak normal atau ada indikasi peretasan. Status koneksi dinyatakan dengan 3 bagian yaitu normal, *warning*, *danger*. Normal untuk koneksi tanpa adanya indikasi peretasan, *warning* untuk koneksi dengan indikasi peretasan tingkat menengah, dan *danger* untuk

koneksi dengan indikasi peretasan tingkat tinggi. Berikut hasil dari nilai *fitness* :

```

maulana/PycharmProjects/guessPassword/venv/bin/python /home/r
['0', '1', '1', '1', '1'] 4 0:00:00.000692 Warning
['1', '1', '1', '0', '1'] 4 0:00:00.002012 Warning
['1', '0', '1', '0', '1'] 3 0:00:00.002546 Normal
['1', '1', '1', '1', '1'] 5 0:00:00.003230 Danger
['1', '1', '1', '1', '1'] 5 0:00:00.003891 Danger

```

Gambar 4 Hasil nilai fitness

Dari gambar diatas terdapat nilai fitness ditunjukkan dengan angka 4, 4, 3, 5, 5. Nilai fitness tersebut masuk pada 3 kategori yaitu 2 koneksi dengan dengan nilai fitness 4 masuk pada status warning, 1 koneksi dengan nilai fitness 3 masuk pada status normal, dan 2 koneksi dengan nilai fitness 5 masuk pada status danger. Status tersebut diperoleh dari nilai fitness dengan ketentuan apabila nilai fitness lebih besar dari 4 maka nilai status adalah danger, apabila nilai fitness lebih besar dari 3 maka nilai status adalah warning, selain dari ketentuan tersebut maka status dapat dinyatakan normal.

IV. SIMPULAN

Hasil penelitian menunjukkan mencari similarity antara log server normal dan log server tidak normal dapat mendeteksi adanya percobaan serangan atau dapat mengetahui adanya koneksi yang tidak normal pada server. Namun tidak efektif terhadap serangan yang belum diketahui sebelumnya.

Untuk meningkatkan kualitas deteksi maka harus adanya penambahan data log tidak normal. Penambahan data log tidak normal tentunya akan mempegaruhi kualitas deteksi.

DAFTAR PUSTAKA

- Arkeman, Y., Seminar, K. B., & Gundawan, H. (2012). Algoritma Genetika Teori dan Aplikasinya untuk Bisnis dan industri (p. 205). PT Penerbit IPB Press.
- Fahmi, I., Fachruddin, R., & Silvia, V. (2019). Analisis E-Business pada Keputusan Kewirausahaan Bisnis Digital (Sebuah Kajian Literatur Studi). In Prosiding Seminar Nasional Kewirausahaan (Vol. 1, No. 1, pp. 71-75).
- Kurniawati, A., Puspitodjati, S., & Rahman, S. (2010). Implementasi Algoritma Jaro-Winkler Distance untuk Membandingkan Kesamaan Dokumen Berbahasa Indonesia. Proceeding, Seminar Ilmiah Nasional Komputer Dan Sistem Intelijen KOMMIT 2008, Depok, Indonesia.
- Mahmudy, F. W. (2013). Algoritma Evolusi . Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya.
- Mahmudy, WF, Marian, RM & Luong, LHS. (2014). Hybrid genetic algorithms for part type selection and machine loading problems with alternative production plans in flexible manufacturing system. (Vol. 8, No. 1, pp. 80-93).
- Rahardjo, B. (2005). Keamanan Sistem Informasi Berbasis Internet. In PT Insan Infonesia.
- Rahmathulla, V.K. Das P. Ramesh, M. & Rajan, R.K. (2007). Growth Rate Pattern and Economic Traits of Silkworm Bombyx mori, L under the influence of folic acid administration. J. Appl. Sci. Environ. Manage. 11(4): 81-84
- Zukhri, Z. 2004. "Penyelesaian Masalah Penugasan dengan Algoritma Genetika". Seminar Nasional Aplikasi Teknologi Informasi, Yogyakarta : Universitas Islam Indonesia.