Coding: Jurnal Komputer dan Aplikasi
Volume 07, No. 01 (2019), hal 120-130 ISSN: 2338-493X

ANALISA *LOG WEB SERVER* UNTUK MENGETAHUI POLA PERILAKU PENGUNJUNG *WEBSITE* MENGGUNAKAN TEKNIK *REGULAR EXPRESSIONS*

[1]Yogi, [2]Ikhwan Ruslianto, [3]Syamsul Bahri
[1,2,3] Jurusan Rekayasa Sistem Komputer, Fakultas MIPA Universitas Tanjungpura
Jl. Prof. Dr. H. Hadari Nawawi, Pontianak
Telp./Fax.: (0561) 577963
e-mail: [1]yogi_bears12@student.untan.ac.id, [2]ikhwanruslianto@siskom.ac.id,
[3] syamsul.bahri@siskom.untan.ac.id.

ABSTRAK

Access log merupakan daftar semua aktifitas pengunjung selama mengakses suatu website. Informasi tersebut akan sangat berguna apabila website mengalami serangan sehingga akan dapat dicari penyebabnya berdasarkan log file yang terdapat pada web server. Namun, untuk mencari informasi yang relevan dari data yang terkait dengan serangan pada log file itu menjadi tugas yang sulit bagi seorang administrator website. Banyaknya permintaan yang dilakukan oleh pengunjung website, akan meningkatkan ukuran rekaman data log yang tersimpan dalam sebuah website. Pada penelitian ini, dibuat sebuah sistem yang berfungsi sebagai alat untuk mengetahui informasi aktivitas pengunjung pada sebuah website menggunakan data access log dengan teknik regular expression. Regular Expression merupakan sebuah bahasa mini untuk mendeskripsikan string atau teks. Regular Expression dapat dipakai untuk mencocokkan sebuah string dengan sebuah pola. Pengujian dilakukan dengan data access log dari tiga website yaitu katakutu.net, berkahbarang.id dan screen6.id. Berdasarkan hasil penelitian, sistem mampu memberikan informasi mengenai pola kunjungan seperti jumlah kunjungan halaman terbanyak, informasi data pengguna dan aktifitas berbahaya pada website. Hasil pengujian pada sistem, mendeteksi 46 serangan Cross Site Scripting, dan 983 serangan Path Traversal dengan total serangan sebanyak 1029 serangan yang merupakan aktifitas berbahaya pada website.

Kata Kunci: *Access Log, Regular Expression, Web Log Analyzer, Web Mining, Vulnerabilities.*

1. PENDAHULUAN

Saat terjadi serangan pada *website,* administrator akan memeriksa *log* untuk mengetahui darimana serangan berasal. Dengan memeriksa satu per satu setiap catatan yang tersimpan pada *log*, maka data-data seorang peretas akan diketahui. Data peretas tersebut, dapat diketahui dengan cara melihat dari alamat IP yang dipakai untuk mengakses *web server*. Namun, catatan yang tersimpan dalam *log*, tentunya berisi catatan seluruh pengunjung yang mengakses *web server* tersebut, akan menjadi tidak efisien apabila metode pencarian data itu diperiksa satu per satu dari banyaknya data yang tersimpan [1]. Semua *request* yang diproses oleh *web server* terekam dalam sebuah *log*, yakni *access log*. *Access log* berisi data informasi pengunjung *website* seperti IP *address*, waktu akses *website*, *request* yang dikirimkan, *web browser* yang digunakan, dan informasi penting lainnya [2]. Informasi ini dapat digunakan untuk proses analisa *website* untuk mengetahui pola perilaku pengunjung *website* untuk meninjau aktivitas berbahaya pada *website* yang dikelola [3].

Salah satu penelitian sebelumya dari Muhammad Dahria yang berjudul "*Analisis Web Server Log Dalam Pencarian Pola Pengunjung Dengan Teknik Association Rules*" membahas penggunaan teknik *Association Rules* untuk menentukan proses analisis *web server log* dalam pencarian pola kunjungan *website* Pemerintah Kota Medan dengan menerapkan metode yang tepat yaitu metode *Association Rule* yang mampu menghasilkan aturan-aturan asosiasi yang menggambarkan pola kunjungan terhadap *website* Pemko Medan [4]. Penelitian dari Muhammad Dahria hanya berfokus pada pola kunjungan *website* dan tidak membahas adanya aktivitas berbahaya pada *website*. Kemudian penelitian dari Benny Nixon yang berjudul "*Pengembangan Program Penyaringan Data Web Log Untuk Analisis Pola Akses Pengunjung Web server*" membahas analisis data *web log*, hasil analisis pengunjung *web server* menampilkan informasi trafik, halaman *website* yang paling banyak dikunjungi, kode

status halaman *website*, *browser* yang digunakan *user*, *user access pattern*, dan *behavior user pattern.* Pengolahan *data log* dilakukan dengan memproses *data log* yang diubah dalam format .csv untuk mempermudah menyimpan data ke dalam *Database* MySQL [5]. Proses pengolahan data pada penelitian tersebut masih belum efisien dan penelitian ini hanya berfokus pada pola kunjungan *website* dan tidak membahas adanya aktivitas berbahaya pada *website*.

Penelitian analisa *log web server* ditujukan untuk mendapatkan pola perilaku pengunjung yang akan diuji coba pada *website* katakutu.net, berkahbarang.id dan screen6.id. Informasi pola kunjungan ini dapat digunakan untuk mengetahui halaman mana saja yang sering diakses, penguna yang mengakses *website*, dan aktifitas berbahaya pada *website*. Untuk mengatasi masalah tersebut pada penelitian ini akan membahas penggunaan Teknik *Regular Expression* untuk mendeteksi adanya pola perilaku pengunjung *website* yang dilakukan oleh *client. Regular Expression* (ekspresi regular, *Regexp*, *Regex*, RE) adalah sebuah bahasa mini untuk mendeskripsikan *string* atau teks [6]. *Regular Expression* dapat dipakai untuk mencocokkan sebuah *string* dengan sebuah pola. Dalam hal ini peneliti akan melakukan analisa pada *log website* yang berfokus untuk mendeteksi adanya serangan pada *website* sehingga para administrator *website* dapat melakukan pencegahan maupun proses perbaikan pada *website* yang dikelola.

Kebutuhan akan suatu sistem untuk mengetahui dan mendeteksi pola perilaku pengunjung terhadap suatu aplikasi *web* sangatlah diperlukan untuk melakukan pengembangan halaman *website* yang lebih baik dimasa yang akan datang. Karena itu, dalam tugas akhir ini akan dibangun suatu sistem yang dapat mengetahui pola pengunjung perilaku pada aplikasi *website* dalam *log* menggunakan Teknik *Regular Expression*. Dalam tugas akhir ini juga akan membahas serangan pada *website*, serangan yang akan dibahas adalah *Cross-Site Scripting* dan *Path Traversal Attack*. Walaupun nilai dari sebuah keamanan tidak ada yang mutlak, harapannya dengan sistem yang telah dibangun nantinya dapat mengurangi peluang seorang peretas untuk dapat meretas aplikasi *website*.

2. LANDASAN TEORI

2.1 *Web Server*

Web Server adalah tempat untuk mendapatkan halaman *web* dan data yang berhubungan dengan *website* yang di buat, sehingga data dapat diakses dan dilihat oleh pengguna [7]. Namun, *web server* dapat mempunyai dua pengertian berbeda, yaitu sebagai bagian dari perangkat keras (*hardware*) maupun sebagai bagian dari perangkat lunak (*software*). Jika merujuk pada *hardware*, *web server* digunakan untuk menyimpan semua data seperti HTML dokumen, gambar, file CSS *stylesheets*, dan file *JavaScript*. Sedangkan pada sisi *software*, fungsi *web server* adalah sebagai pusat kontrol untuk memproses permintaan yang diterima dari *browser*.

Gambar 1. Cara Kerja *Web Server*

Ada tiga macam *log* yang harus diperhatikan dalam memonitor kegiatan *web server* yaitu sebagai berikut:

1. *Access Log*

Access log adalah *file* yang berfungsi untuk mencatat semua akses yang dilakukan terhadap *web server*.

2. *Server Log*

Server log adalah *file* yang berfungsi untuk mencatat kejadian-kejadian tertentu pada *web server*. Tetapi, *file* ini hanya diperiksa jika ternyata *webserver* ini terjadi kesalahan.

3. *Error Log*

Error log adalah *file* yang berfungsi untuk mencatat setiap kesalahan yang terjadi pada web server, apakah kesalahan terjadi pada file konfigurasi maupun kesalahan pada pembuatan *website*.

2.2 Access Log

Access Log adalah data yang mencatat tindakan yang terjadi pada sebuah *web*site. *Data log* ini berada di komputer *web server*, *log* ini dapat dilakukan untuk melacak dan menganalisis perilaku pengguna yang mengunjungi situs *web* [8]. *Log Web* berisi informasi tentang *Admin IP Address*, *Admin Name*, URL, Permintaan Akses, *Time Stamp*, kode kesalahan dll [9]. Berikut adalah contoh format pada *web log* dari

https://httpd.apache.org/docs/2.4/logs.HTML:

```
"%h %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{Admin-agent}i\""
```

Keterangan:

- %h: *Remote IP address* atau *domain name*.
- %l: Identitas admin yang ditentukan dari *identd*.
- %u: merupakan nama *admin* yang ditentukan oleh autentikasi HTTP.
- %t: Tanggal dan waktu akses pengunjung.
- \"%r\": Mode permintaan(Request) : *GET, POST*, atau *method HEAD* dari CGI (*Common Gateway Interface*)
- %>s: Kode status saat *client* mengunjungi suatu halaman *website*. Kode 200 adalah "OK", jika 404 maka "Tidak Ditemukan atau "*Not Found*".
- %b: Kapasitas dokumen yang ditransfer (Bytes).
- \"%{*Referer*}i\": merupakan halaman yang terkait dengan URL ini.
- \"%{Admin-*agent*}i\": merupakan informasi mengenai *browser* yang digunakan pengunjung.

Format *web log* akan di modifikasi agar menampilkan informasi yang benar-benar dibutuhkan oleh administrator *web*. Adapun tahapan yang diperlukan selama proses modifikasi *data log web* diantaranya adalah *raw web log data*, data *cleaning*, identifikasi *user*, identifikasi *session*, dan *Database of clean log* [5].

2.2.1 *Raw Web Log Data*

Raw Web Log Data adalah proses penyiapan data *web log* yang ada pada *web server*. Data yang ada pada *web log* tidak akan langsung digunakan dalam proses analisis karena banyak informasi yang tidak relevan untuk mendapatkan data pola akses pengunjung dari *web*site.

2.2.2 Data *Cleaning*

Proses ini dilakukan untuk menyaring informasi yang tidak relevan pada *log web*, sehingga *log web* hanya menyimpan data yang dibutuhkan saja untuk proses selanjutnya. [10]. Namun dalam penelitian ini peneliti tidak melakukan modifikasi data *cleaning*, agar semua aktifitas pada data *log web* dapat dianalisa sebagai bagian pola perilaku pengunjung pada *website*.

2.2.3 Identifikasi *User*

Beberapa *user* mungkin menggunakan komputer (*host*) yang sama, maka perlu dilakukan proses identifiaksi admin. Identifikasi Admin adalah Proses identifikasi setiap admin yang mengakses *website*.

2.2.4 Identifikasi *Session*

Setelah admin diidentifikasi, halaman yang diakses pun harus dibagi kedalam sesi tertentu, umumnya berdasarkan waktu tertentu agar didapatkan sesi yang tunggal untuk setiap admin. Tujuan dari *session identification* adalah untuk menemukan pola akses setiap pengguna dan halaman yang sering diakses. Metode yang paling sederhana adalah menggunakan *timeout*, dimana jika waktu antara permintaan halaman melebihi batas tertentu, dianggap pengguna memulai *session* baru. Banyak produk komersial menggunakan waktu 30 menit sebagai nilai *default timeout*.

2.2.5 *Database Clean of Log*

Proses *Database of Clean Log* merupakan tahapan akhir setelah tahapan *raw web log data*, data *cleaning*, identifikasi *admin*, dan identifikasi *session*. Pada tahapan ini data *web log* sudah siap di proses dan digunakan untuk menentukan pola akses pengunjung *web server*.

2.3 *Regular Expressions*

Regular Expression (ekspresi *regular, Regexp, Regex, RE*) adalah sebuah bahasa mini untuk mendeskripsikan *string* atau teks. *Regex* dapat dipakai untuk mencocokkan sebuah string dengan sebuah pola. Yang membedakan *Regex* dari *string* biasa adalah terdapatnya karakter-karakter khusus yang disebut karakter meta (*metacharacters*). Karakter-karakter ini tidak akan dicocokkan secara literal dengan karakter itu sendiri, tapi mewakili sekelompok karakter lain atau pola khusus tertentu [11].

Regex biasa digunakan untuk mencocokkan dan mencari pola dalam teks, dari pola yang sederhana hingga yang sangat kompleks. Berikut beberapa contoh karakter *Regex* yang digunakan menurut Michael Fitzgerald dalam buku yang berjudul *Introducing Regular Expressions* [12] pada Tabel 1 yaitu:

Tabel 1. *Metacharacter Regex*

No	Metacharacter	Keterangan

1	/	digunakan untuk memulai atau mengakhiri baris *Regex*. Notasi awal atau akhir *Regex*
2	\	digunakan dalam *Regex*, yang mana jika disandingkan dengan sebuah karakter, maka karakter tersebut akan dicocokkan secara literal sebagai karakter itu sendiri
3	^	*Regex* yang mencocokkan karakter *null* pada awal baris
4	( )	digunakan untuk membuat *Regex* yang berada di dalamnya menjadi sebuah grup atau mencocokkan teks yang sama dengan *Regex* aslinya
5	[ ]	digunakan untuk menunjukkan kelas karakter (*character class*) yang mencocokan salah satu karakter pada set karakter (*character set*)
6	[< string >]	*Regex* yang mencocokkan apa pun dari karakter pada *string* dan tidak ada lainnya
7	[^< string >]	*Regex* yang mencocokkan karakter apa pun kecuali baris baru dan karakter dari *string* itu sendiri
8	+	*Regex* yang mencocokkan *string* atau karakter itu sendiri sebanyak sekali atau lebih
9	“	karakter yang dibaca oleh *Regex* sebagai karakter itu sendiri yaitu tanda petik
10	.	*Regex* yang bertindak sebagai karakter khusus dan akan mencocokkan karakter apa pun, kecuali baris baru
11	-	karakter literal untuk mencocokkan tanda hubung
12	$	*Regex* yang yang mencocokkan karakter *null* sebelum karakter baris baru, biasanya berada pada akhir baris
13	*	*Regex* yang mencocokkan angka apa pun (termasuk nol) dan mencocokkannya nol kali atau lebih
14	\S	Pola *Regex* untuk mencari baris yang tidak diawali spasi atau tab
15	\s	Pola *Regex* untuk mencari baris yang diawali spasi atau tab
16	\d	Pola *Regex* untuk mencari baris karakter angka (0-9), setara dengan [0-9]
17	\D	Pola *Regex* untuk mencari baris bukan karakter angka (0-9), setara dengan [^0-9]

2.4 *Web Vulnerabilities Attack*

Web Vulnerability Attack merupakan jenis serangan terhadap aplikasi *web* berdasarkan kerentanan yang terdapat pada aplikasi. Pada penelitian ini, ada 4 macam serangan yang akan menjadi bahasan. Serangan-serangan tersebut adalah sebagai berikut:

2.4.1 *Cross Site Scripting*

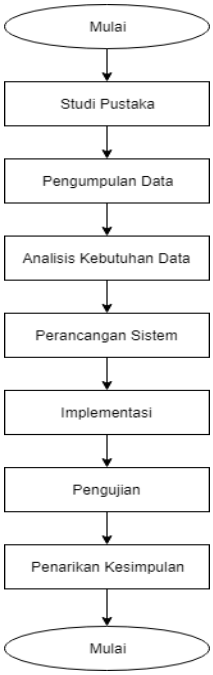
Cross-Site Scripting (XSS) adalah serangan yang memungkinkan penyerang mengeksekusi skrip di *browser* korban untuk membajak sesi pengguna dan mencuri miliknya, merusak situs *web*, menyisipkan konten yang tidak ramah, mengalihkan pengguna, membajak peramban pengguna menggunakan *malware*, dll [13].

2.4.2 *Path Traversal*

Path Traversal Attack merupakan sebuah serangan yang mengakses berkas-berkas yang terdapat pada penyimpanan lokal *server* yang seharusnya berkas tersebut tidak dapat diakses melalui aplikasi *web*. Celah ini biasanya ada pada aplikasi yang menyisipkan *script* dari berkas lain pada suatu halaman atau berkas [14].

3. METODE PENELITIAN

Pada bab ini dibahas tentang alur penelitian yang dilakukan meliputi diagram alir penelitian seperti pada Gambar 3.2



Gambar 2. *Diagram* Alir Penelitian

3.1 Studi Pustaka

Studi pustaka bertujuan untuk mempelajari teori-teori yang berhubungan dengan penerapan *Regular Expression* pada Analisa serangan pada *log web*. Studi literatur dilakukan dengan cara membaca buku-buku baik lokal maupun buku-buku internasional yang diterjemahan dan jurnal penelitian yang terkait serta artikel dari *internet*. Hasil dari pelaksanaan studi literatur ini adalah dengan adanya pengetahuan mengenai metode *Regular Expression* untuk mendeteksi serangan pada *log web* berdasarkan parameter atau metode serangan. Serangan yang akan dianalisa adalah *Cross-Site Scripting* dan *Path Traversal Attack.*

3.2 Metode Pengumpulan Data

Metode pengumpulan data ini meliputi data-data yang dapat menunjang penelitian yang berkaitan dengan informasi *data log* web untuk proses Analisa serangan. *Data log* web merupakan data yang diperoleh dari *web server*. Data pada penelitian ini berupa data format *log web* seperti *IP Address, date, time, request mode, kode status, byte, refferer* dan *admin agent* digunakan untuk mengidentifikasi *admin* dan *session*. Format *web log* akan di modifikasi agar menampilkan informasi yang dibutuhkan oleh administrator *web*.

3.3 Analisis Kebutuhan

Pada tahap ini peneliti menganalisis kebutuhan perangkat lunak dan perangkat keras. Kebutuhan perangkat keras yang dibutuhkan dalam mengerjakan penelitian tugas akhir ini meliputi: Prosesor AMD A8-7410, RAM 4 GB Hardisk 240 GB.

Kebutuhan perangkat lunak yang digunakan untuk membangun sistem ini antara lain: *windows 10, balsamiq mockup*, MySQL, *Google Chrome*, *Sublime*, *Notepad++*, XAMPP, dan *draw.io*.

3.4 Perancangan Sistem

Perancangan sistem dibuat berdasarkan penelitian yang akan dilakukan. Pada skripsi ini, tahap perancangan dibagi menjadi dua tahapan, yaitu analisis kebutuhan dan perancangan sistem. Sistem ini akan memiliki 6 menu yaitu halaman *Dashboard, Most Visited Pages, Users, Unusual Behavior,* Histori *Log,* dan Info Grafik. Didalam menu *Unusual Behavior* terdapat 4 tipe serangan yaitu, *Cross-Site Scripting* dan *Path Traversal*. Perancangan sistem bertujuan agar aplikasi yang dibuat lebih terarah dan sebagai acuan dalam membuat aplikasi ke dalam kode program.

3.5 Implementasi

Implementasi merupakan tahap pembangunan sistem yang berdasarkan hasil analisis kebutuhan dan perancangan sistem yang telah dilakukan. Sistem akan dibangun dimulai dari perancangan antarmuka dan pembentukan sistem basis data.

3.6 Pengujian

Pada tahap ini peneliti melakukan pengujian sistem untuk melihat sejauh mana sistem berfungsi dan berjalan dengan baik, uji coba dilakukan dengan melihat aktivitas menu yang dibuat dan melakukan serangan pada *web server* yang akan diambil *data log* untuk diujikan kedalam sistem. Serangan yang dilakukan adalah *Cross Site Scripting* dan *Path Traversal Attack.*

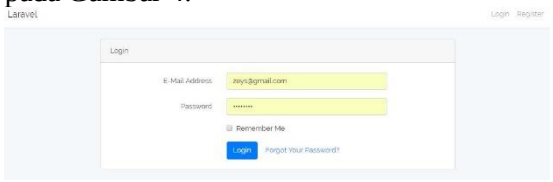
4. HASIL DAN PEMBAHASAN

4.1 Impelementasi Sistem

Pada halaman utama terdapat menu *login*, *register*, *contact* dan *about.* Halaman ini merupakan halaman pertama untuk masuk kedalam sistem aplikasi. Tampilan halaman utama dapat dilihat pada Gambar 3.

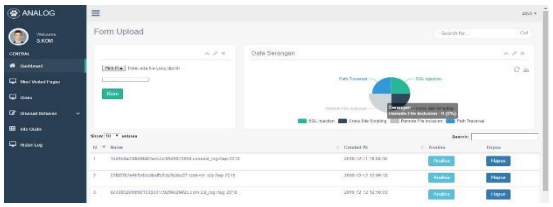
Gambar 3. Halaman Utama

Pada halaman *login* terdapat masukan alamat *email* dan memiliki tombol konfirmasi *login*. Tampilan halaman utama dapat dilihat pada Gambar 4.



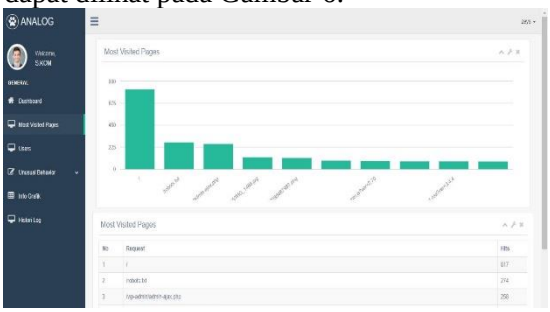
Gambar 4. Halaman *Login*

Pada halaman *dashboard* terdapat tombol pilih *file* sebagai tempat untuk unggah akses *log*. *Log* yang sudah selesai di unggah akan tampil di halaman *dashboard*, disini admin dapat melakukan analisa *log* maupun hapus *data log*. Selain itu terdapat juga tampilan informasi data serangan yang ada pada akses *log*. Tampilan halaman *dashboard* dapat dilihat pada Gambar 5.



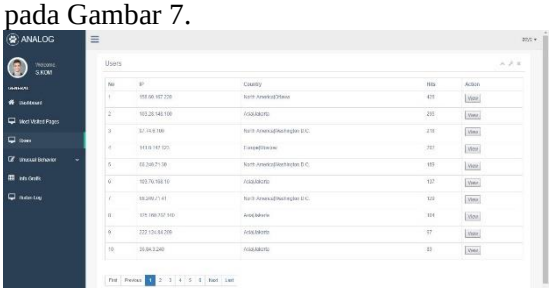
Gambar 5. Halaman *Dashboard*

Pada halaman *most visited pages* terdapat informasi data akses *log* yang menampilkan *request* yang paling sering diakses pada *website*. Tampilan halaman *most visited pages* dapat dilihat pada Gambar 6.



Gambar 6. Halaman *Dashboard*

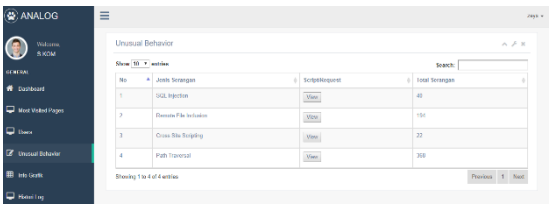
Pada halaman *users* terdapat informasi data akses *log* yang menampilkan data pengguna seperti *ip*, *country*, *hits*, dan tombol *view*. Tampilan halaman *users* dapat dilihat pada Gambar 7.



Gambar 7. Halaman *Users*

Gambar 8 merupakan lanjutan dari halaman *users* saat admin menekan tombol *view*, sistem akan menampilkan informasi data *date*, *time*, dan *request* dari satu ip pengguna.

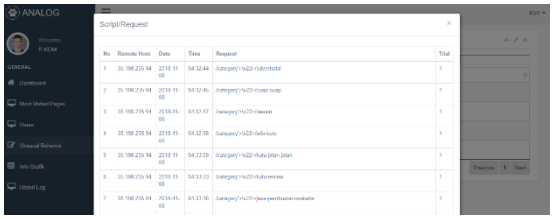
Gambar 8. Halaman *Users Info*



Gambar 9. Halaman *Unusual Behavior*

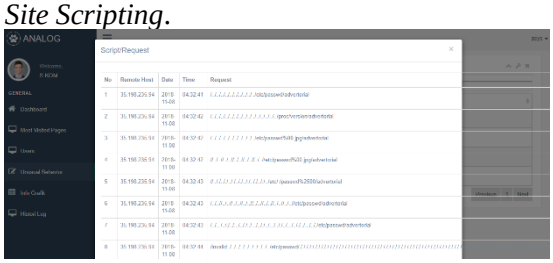
Gambar 9 merupakan halaman *unusual behavior* yang berisi informasi serangan yang ada pada *data log* yaitu *Cross Site Scripting*, dan *Path Traversal.* Tombol *view* untuk melihat *script*/*request* yang dilakukan pengguna dan total serangan untuk mengetahui

informasi percobaan serangan ke dalam *website*.



Gambar 10. Halaman *View Cross Site Scripting*

Gambar 10 menampilkan implementasi halaman *Cross Site Scripting* yang merupakan salah satu pilihan dugaan serangan pada halaman *unusual behavior*. Pada halaman ini terdapat kolom *date* dan *time* yang menunjukkan waktu serangan, kolom *IP address* yang menunjukkan siapa yang menyerang *website*, dan kolom *query/script* yang menunjukkan kode yang digunakan penyerang dalam melancarkan serangan *Cross Site Scripting*.



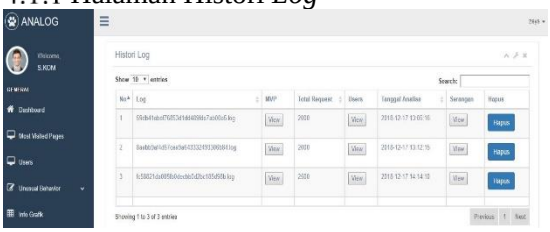
Gambar 11. Halaman *View Path Traversal*

Gambar 11 menampilkan implementasi halaman *Path Traversal* yang merupakan salah satu pilihan dugaan serangan pada halaman *unusual behavior*. Pada halaman ini terdapat kolom *date* dan *time* yang menunjukkan waktu serangan, kolom *IP address* yang menunjukkan siapa yang menyerang *website*, dan kolom *query/script* yang menunjukkan kode yang digunakan penyerang dalam melancarkan serangan *Path Traversal*.

Gambar 12. Halaman Info Grafik

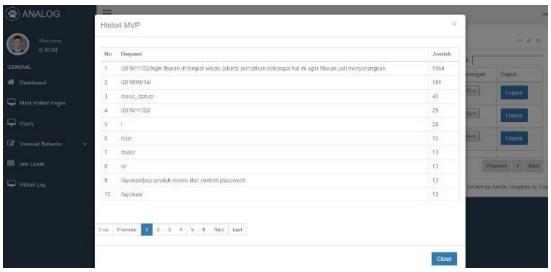
Gambar 12 merupakan halaman yang menampilkan visualisasi data akses *log* dalam bentuk grafik. Data yang ditampilkan merupakan halaman *website* yang diantaranya adalah *Most Visited Pages* untuk melihat data kunjungan halaman terbanyak, *Unusual Behavior* untuk melihat data serangan pada *website*, dan Data *IP Client* terbanyak yang mengunjungi *website*.

4.1.1 Halaman Histori *Log*



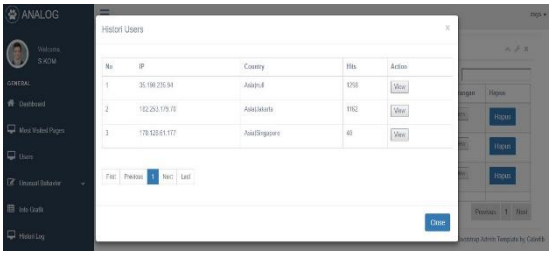
Gambar 13. Halaman Histori *Log*

Gambar 13 merupakan halaman yang menampilkan informasi data yang sudah dianalisa. Data yang sudah terhapus di halaman *dashboard* masih tersimpan di halaman histori *log* dan masih bisa dilihat data-data yang sudah dianalisa. Data yang ditampilkan di halaman histori *log* adalah informasi nama *log*, *view most visited pages* untuk melihat histori data kunjungan halaman terbanyak, total *request*, *view users* untuk melihat data histori IP *Client* terbanyak yang mengunjungi *website*, tanggal analisa, *view* serangan untuk melihat histori data serangan pada *website*, dan tombol hapus.



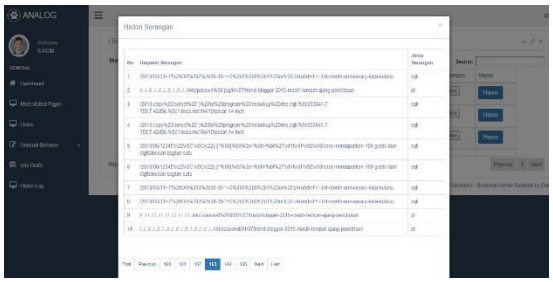
Gambar 14. Halaman *View* Histori *Most Visited Pages*

Gambar 14 merupakan halaman yang menampilkan informasi data histori halaman *most visited pages*. Data yang ditampilkan berupa informasi *request* dan jumlah kunjungan yang diakses.



Gambar 15. Halaman *View* Histori *Users*

Gambar 15 merupakan halaman yang menampilkan informasi data histori halaman *users*. Data yang ditampilkan berupa informasi *ip*, *country*, *hits*, dan tombol *view*.



Gambar 16. Halaman *View* Histori *Users*

Gambar 16 merupakan halaman yang menampilkan informasi data histori halaman serangan dari halaman *unusual behavior* yang sudah dianalisa. Data yang ditampilkan berupa informasi *query request* serangan dan jenis serangan.

4.2 Tahapan Proses Analisa Pola Perilaku Pengunjung

Proses analisa dimulai dengan membaca file web log yang telah melalui tahapan raw web log data, data cleaning, identifikasi user, identifikasi session, dan Database of clean log. Proses dilanjutkan dengan menganalisis baris data untuk menentukan apakah baris data tersebut mengandung data yang merepresentasikan fungsi tertentu.

Dari analisis diatas diharapkan pengelola *website* dapat mengetahui pola akses pengunjung *web server* untuk perbaikan kinerja *website*, berkaitan dengan:

1. Halaman yang sering diakses pengguna *website*

Informasi yang terdapat pada setiap halaman *web* menjadi hal yang penting untuk setiap pengunjung suatu *website*. Seberapa sering suatu halaman *web* diakses oleh pengunjung dapat menjadi perhatian bagi pengelola *website*.

2. Data *users* yang mengakses *website*

Data pengguna pada *website* akan memberikan informasi kepada pengelola *website* untuk mengetahui aktifitas apa saja yang dilakukan *user*. Informasi yang didapat berupa ip, negara, kota, *hits* atau jumlah kunjungan, dan *request* yang diakses oleh pengguna *website*.

3. Aktifitas berbahaya pada *website*

Saat terjadi serangan, pengelola *website* dapat mengetahui informasi *user* yang melakukan dan serangan apa yang dilakukan *user*. Dalam penelitian ini serangan yang dideteksi adalah *Cross Site Scripting*, dan *Path Traversal*.

4.2.1 Pengujian Data

Pengujian data dilakukan dengan menguji 30 data akses *log* yang terbagi dalam 3 *website* berbeda yaitu katakutu.net sebanyak 19 data, berkahbarang.id sebanyak 8 data, dan screen6.id sebanyak 3 data. Pengujian dilakukan dengan mengunggah masing-masing akses *log* untuk dianalisa. Penjelasan lebih lanjut akan dijabarkan pada setiap pengujian data serangan dalam menu *most visited pages*, *users*, dan *unusual behaviour*.

1. *Most Visited Pages*

Tabel 2. *Most Visited Pages* katakutu.net

No	Request	Hits
1	/basic_status/	33966
2	/	4668
3	/robots.txt	352
4	/wp-login.php	285
5	/feed/	149

Tabel 2 merupakan halaman *website* yang paling banyak dikunjungi dari data log katakutu.net. Halaman yang paling banyak dikunjungi adalah /basic_status/ dengan total 33966 kunjungan dan halaman utama "/" dengan total 4668 kunjungan.

Tabel 3. *Most Visited Pages* berkahbarang.id

No	Request	Hits
1	/	1630
2	/robots.txt	646
3	/wp-admin/admin-ajax.php	348
4	/wp-content/uploads/2018/05 /IMG_1488.jpg	198
5	/wp-content/uploads/2017/10 /LOGObiru- e1509098487983. png	167

Tabel 3 merupakan halaman website yang paling banyak dikunjungi dari data log katakutu.net. Halaman yang paling banyak dikunjungi adalah halaman utama "/" dengan total 1630 kunjungan.

Tabel 4. *Most Visited Pages* berkahbarang.id

No	Request	Hits
1	/	250

2	/wp-includes/js/thickbox/thickbox .css?ver=4.6.3	31
3	/wp-content/uploads/2017/02/tim eline-dota.png	31
4	/wp-content/plugins/widget-countdown/includes/style/style.css?ver=4.6.3	31
5	/wp-content/plugins/contactform-/includes/css/styles.css?ver=4.6.1	31

Tabel 4 merupakan halaman website yang paling banyak dikunjungi dari data log screen6.id. Halaman yang paling banyak dikunjungi adalah halaman utama "/" dengan total 250 kunjungan.

2. *Users*

Tabel 5. *Users* katakutu.net

No	IP	Country	Hits
1	178.128.61.177	Asia|Singapore	46255
2	35.198.235.94	Asia|Null	5575
3	95.216.20.167	Europe|Tuusula	300
4	194.187.170.114	Europe|Null	160
5	66.249.69.231	North America|Null	93

Tabel 5 merupakan data users dari *website* katakutu.net. IP 178.128.61.177 merupakan data pengguna yang paling banyak melakukan kunjungan pada *website* katakutu.net dengan total 46255 kunjungan dan berasal dari kota Singapore.

Tabel 6. *Users* berkahbarang.id

No	IP	Country	Hits
1	158.69.167.220	North America|Ottawa	856
2	141.8.142.123	Europe|Moscow	462
3	97.74.6.100	North America|Washington D.C	436
4	66.249.71.39	North America|Washington D.C	378
5	66.249.71.41	North America|Washington D.C	258

Tabel 6 merupakan data users dari *website* berkahbarang.id. IP 158.69.167.220 merupakan data pengguna yang paling banyak melakukan kunjungan pada website berkahbarang.id dengan total 856 kunjungan dan berasal dari kota Ottawa.

Tabel 7. *Users* screen6.id

No	IP	Country	Hits
1	180.248.171.30	Asia|Pontianak	100
2	36.77.204.62	Asia|Jakarta	90
3	36.82.181.56	Asia|Pontianak	87
4	180.248.164.233	Asia|Pontianak	82
5	125.160.84.201	Asia|Pontianak	80

Tabel 7 merupakan data users dari website screen6.id. IP 180.248.171.30 merupakan data pengguna yang paling banyak melakukan kunjungan pada website screen6.id dengan total 100 kunjungan dan berasal dari kota Pontianak.

3. Unusual Behavior

Tabel 8. Hasil Pengujian *Unusual Behavior*

No	Website	Serangan	
		XSS	PT
1	Katakutu.net	46	983
2	Berkahbarang.id	0	0
3	Screen6.id	0	0
Total		46	983

Tabel 8 merupakan hasil pengujian pada menu unusual behaviour untuk mengetahui aktifitas berbahaya pada *website*. Dari tiga puluh data akses *log* yang sudah dianalisa, ditemukan sebanyak 46 serangan *Cross Site Scripting*, dan 983 serangan *Path Traversal*.

4.3 Pembahasan

Analisa dilakukan dengan mengunggah data log satu per satu dan sistem akan mengolah dan memecah data log, saat proses selesai maka sistem akan menyimpan data di dalam basis data. Setelah itu administrator *website* dapat memilih dan menekan tombol analisa pada data log. Pengujian dilakukan dengan 30 data akses *log* yang terbagi dalam 3 website berbeda yaitu katakutu.net sebanyak 19 data, berkahbarang.id sebanyak 8 data, dan screen6.id sebanyak 3 data.

Berdasarkan hasil pengujian pada website katakutu.net dapat dilihat pola perilaku pengunjung berdasarkan halaman yang paling sering dikunjungi yaitu /basic_status/ dan

halaman utama "/" dengan masing-masing total kunjungan sebanyak 33966 dan 4668. Halaman /basic_status/ merupakan request/permintaan dari sebuah server untuk mengetahui kondisi webserver pengguna secara realtime. Dengan mengabaikan request aktifitas web server dapat dilihat halaman yang sering dikunjungi merupakan halaman "/2016/11/22/10-vga-card-murah-untuk-gaming-dan-desain-grafis" dengan total kunjungan sebanyak 144 kunjungan. Dari informasi data users jumlah kunjungan terbanyak merupakan pengguna IP 178.128.61.177 dengan total kunjungan 46255 yang berasal dari kota Singapore. Data waktu kunjungan website berdasarkan waktu akses didapatkan hasil dengan frekuensi tertinggi terjadi pada pukul 11.00 – 11.59 dan pukul 07.00 – 07.59. Kemudian untuk aktifitas berbahaya sistem mendeteksi 46 serangan Cross Site Scripting dan 983 serangan Path Traversal menggunakan Teknik Regular Expressions.

Hasil pengujian pada website berkahbarang.id dapat dilihat pola perilaku pengunjung berdasarkan halaman yang paling sering dikunjungi yaitu halaman utama "/" dengan total 1630 kunjungan. Dari informasi data *users* jumlah kunjungan terbanyak merupakan pengguna IP 158.69.167.220 dengan total 856 kunjungan dan berasal dari kota Ottawa. Data waktu kunjungan *website* berdasarkan waktu akses didapatkan hasil dengan frekuensi tertinggi terjadi pada pukul 15.00 – 15.59 dan pukul 18.00 – 18.59. Kemudian untuk aktifitas berbahaya sistem tidak mendeteksi adanya serangan.

Hasil pengujian pada *website* screen6.id dapat dilihat pola perilaku pengunjung berdasarkan halaman yang paling sering dikunjungi yaitu halaman utama "/" dengan total 250 kunjungan. Dari informasi data *users* jumlah kunjungan terbanyak merupakan pengguna IP 180.248.171.30 dengan total 100 kunjungan dan berasal dari kota Pontianak. Data waktu kunjungan *website* berdasarkan waktu akses didapatkan hasil dengan frekuensi tertinggi terjadi pada pukul pukul 21.00 – 21.59 dan pukul 23.00 – 23.59. Kemudian untuk aktifitas berbahaya sistem tidak mendeteksi adanya serangan.

Berdasarkan pola tingkah laku *user* tersebut pengelola *website* dapat mempergunakan sebagai acuan untuk memantau kinerja *website* terhadap kemungkinan adanya gangguan, sehingga dapat memberikan kepuasan kepada pengunjung *website.*

5. KESIMPULAN

Berdasarkan hasil perancangan, implementasi, dan pengujian yang dilakukan, maka diambil kesimpulan sebagai berikut:

1. Data yang sudah selesai di unggah akan di analisa oleh sistem dimulai dari halaman *most visited pages* untuk mengetahui halaman yang paling sering diakses oleh pengunjung, halaman *users* untuk mengetahui data pengguna yang mengakses seperti informasi *request,* negara, dan total kunjungan pengguna berdasarkan ip tersebut, halaman *unusual behavior* untuk mengetahui serangan yang ada pada *website*, halaman info grafik untuk mengetahui informasi data dalam bentuk grafik, dan histori *log* untuk mengetahui *data log* yang sudah dianalisa.
2. Hasil analisa dari *website* katakutu.net halaman yang paling sering dikunjungi adalah /basic_status/ dengan total 33966 kunjungan, IP 178.128.61.177 menjadi data pengguna yang paling sering mengunjungi *website* dengan total 46255, frekuensi akses tertinggi terjadi pada pukul 11.00 – 11.59 dan pukul 07.00 – 07.59. Analisa dari *website* berkahbarang.id halaman yang paling sering dikunjungi adalah halaman utama "/" dengan total 1630 kunjungan, IP 158.69.167.220 menjadi data pengguna yang paling sering mengunjungi *website* dengan total 856, frekuensi akses tertinggi terjadi pada pukul 15.00 – 15.59 dan pukul 18.00 – 18.59. Kemudian analisa dari website screen6.id halaman yang paling sering dikunjungi adalah halaman utama "/" dengan total 250 kunjungan, IP 180.248.171.30 dengan total 100 kunjungan, frekuensi tertinggi terjadi pada pukul 21.00 – 21.59 dan pukul 23.00 – 23.59. Dan aktifitas berbahaya ditemukan pada *website* katakutu.net dengan pola serangan *Cross Site Scripting* dan *Path Traversal*. Sedangkan pada *website* berkahbarang.id dan screen6.id tidak ditemukan adanya aktifitas berbahaya pada *website.*
3. Hasil pengujian analisa *log* mendeteksi 46 serangan *Cross Site Scripting* dan 983

serangan *Path Traversal*, total serangan yang didapat sebanyak 1029 serangan pada *website* katakutu.net. Untuk analisa *log* pada *website* screen6.id dan berkahbarang.id tidak ditemukan serangan pada *website*.

6. SARAN

Berdasarkan hasil yang diperoleh dalam penelitian ini, maka penulis memberikan beberapa saran yang perlu menjadi bahan pertimbangan untuk penelitian selanjutnya yaitu:

1. Saat sistem sedang melakukan proses memecah data akses *log* pada basis data, sistem cenderung lama dalam pemrosesan. Maka dibutuhkan *source code* yang sederhana untuk mempercepat komputasi.
2. Pengembangan sistem dapat dilakukan dengan menambah variasi tipe serangan yang dapat dideteksi oleh sistem.
3. Pengembangan sistem yang sudah dibagun dapat dilakukan dengan membuat versi *real time* dan fitur *Artificial Intelligence* dari sistem analisa *log website*.

DAFTAR PUSTAKA

[1] T. A. Cahyanto, "Investigasi Forensika Pada Log Web Server untuk Menemukan Bukti Digital Terkait dengan Serangan Menggunakan Metode Hidden Markov Models," Fakultas Teknologi Industri UII, Yogyakarta, 2014.

[2] L. J. Grace, "ANALYSIS OF WEB LOGS AND WEB USER IN WEB MINING," International Journal of Network Security & Its Applications (IJNSA), Chennai, 2011.

[3] N. Goel, "Analyzing Users Behavior from Web Access Logs using Automated Log Analyzer Tool," International Journal of Computer Applications (0975 – 8887), Banasthali Vidyapith, 2013.

[4] M. Dahria, "ANALISIS WEB SERVER LOG DALAM PENCARIAN POLA PENGUNJUNG WEB DENGAN TEKNIK ASSOCIATION RULES," Jurnal SAINTIKOM Vol. 13, No. 3, Medan, 2014.

[5] B. Nixon, "Pengembangan Program Penyaringan Data Weblog Untuk Analisis Pola Akses Pengunjung Webserver," Fakultas Teknik Universitas Indonesia, Jakarta, 2010.

[6] S. Haryanto, "Regular Expression," in *Regex: Kumpulan Resep Pemrograman*, Jakarta, PC Media, 2004, p. 188.

[7] T. Agus, "Membuat Web Server Menggunakan Dinamic Domain Name System Pada IP Dinamis," *Jurnal Teknologi Informasi & Komunikasi Digital Zone*, pp. 1-10, 2016.

[8] N. Goel, "Analyzing Users Behavior from Web Access Logs using Automated Log Analyzer Tool," *International Journal of Computer Applications (0975 – 8887)*, pp. 29 - 33, 2013.

[9] D. Sahu, "Detecting Users Behavior from Web Access Logs with Automated Log Analyzer Tool," *International Journal of Computer Science and Information Technologies (IJCSIT)*, pp. 5106-5109, 2014.

[10] K. R., "Identifying User Behavior by Analyzing Web Server Access Log File," *IJCSNS International Journal of Computer Science and Network Security*, pp. 327 - 332, 2009.

[11] S. Haryanto, "Regex: Kumpulan Resep Pemrograman," p. 188, 2004.

[12] M. Fitzgerald, Introducing Regular Expressions, Sabostopol: O'Reilly Media, 2012.

[13] T. S. Mule, "Intrusion Protection against SQL Injection And Cross Site Scripting Attacks Using a Reverse Proxy," Pune: International Journal of Computer Science and Information Technologies, 2014.

[14] F. Rahmat, "Sistem Pendeteksi dan Pencegah Peretasan Terhadap Aplikasi Berbasis Web dengan Teknik Web Application Firewall (WAF)," *JURNAL TEKNIK POMITS Vol. 2, No. 1*, pp. 2301-9271, 2014.