BAB 1
PENDAHULUAN

1.1 Latar Belakang

Website merupakan salah satu aplikasi populer bagi pengguna internet yang bersifat publik. Oleh karena itu, *website* menjadi salah satu pilihan bagi *user* dalam membantu pekerjaannya sehari-hari dengan penggunaan yang mudah serta bisa diakses dimanapun dan kapanpun. Namun, dikarenakan sifatnya yang publik *website* sering mengalami serangan. Sehingga, menyebabkan suatu kerusakan pada *web server* yang mengelolah *website* tersebut. *Web server* yang memiliki tingkat keamanan yang lemah selalu menjadi sasaran yang tepat bagi para *attacker* saat menyerang *web server*.

Ancaman-ancaman pada *website* yang terjadi pada tahun 2017 sudah didata oleh OWASP (*Open Web Application Security Project*) dan sudah tercatat pada OWASP Top 10 Security – 2017. Pada OWASP Top 10 Security, terdapat beberapa ancaman dan tingkat resiko dari dampak serangan yang telah diklasifikasikan oleh OWASP. Tingkat ancaman yang diberi nilai sudah dihitung dengan kalkulator khusus dari NIST (*National Institute of Standards and Technology*) yang disebut CVSS (*Common Vunlnerability Scoring System*) dengan rentang *score* 0.0 sampai 10.0. Ancaman tersebut telah digambarkan pada Tabel 1.1.

Tabel 1.1 Klasifikasi Ancaman pada tahun 2017

No	Ancaman	CVSS Score	Persentase yang Sering Terjadi
1	Injection	8.0	35%
2	Broken Authentication	7.0	74%
3	Sensitive Data Exposure	7.0	28%
4	XML External Entities (XXE)	7.0	2%
5	Broken Access Control	6.0	53%
6	Security Misconfiguration	6.0	79%

7	Cross-site Scripting (XSS)	6.0	77%
8	Insecure Deserialization	5.0	2%
9	Using Component with Known Vulnerabilities	4.7	28%
10	Insufficient Logging & Monitoring	4.0	2%

Sumber: https://www.owasp.org/images/0/0a/OWASP_Top_10_2017_GM_%28en%29.pdf

Dengan adanya pengelompokan ancaman yang dilakukan oleh OWASP pada tahun 2017, menyebabkan serangan menjadi suatu permasalah terbesar bagi *user* dan pengelola *website*. Hal tersebut terjadi karena pada setiap tahunnya terdapat data baru mengenai tingkat ancaman yang terjadi sehari-hari (OWASP, 2017).

Dengan adanya ancaman tersebut, maka seorang *sysadmin* harus melakukan penjagaan yang ketat dalam mengamankan *server* khususnya *web server*. Namun, segala kegiatan pada *web server* khususnya kegiatan penyerangan terhadap *website* telah dicatat pada *log* dari *web server*, macam-macam *log* yang tercatat ialah *access log* dan *error log.*

Access log berisi informasi-informasi akses terhadap website seperti *IP address client*, waktu *client* mengakses *website*, *request* yang dikirimkan oleh *client*, *web browser* yang digunakan, merekam aktivitas pengguna, melacak upaya otentikasi, serta *file-file* yang ada pada HTTP *Service*(Joshila Grace, Maheswari and Nagamalai, 2011). *Error log* berisi informasi kegagalan suatu request dari *user / service* yang ada pada *website*, khususnya yang terjadi pada *web server.* Informasi yang ada pada *error log* ialah waktu *client* mengakses *website*, IP *Address client,* dan kegiatan *error* yang dilakukan *client* (Kabir, 2010). *Log* yang merupakan *record* dari seluruh kegiatan *service* akan berisi sebuah data. Jika *website* tersebut terdapat banyak *user* yang mengakses, maka data *log* semakin banyak dan *sysadmin* akan sulit membaca *log* tersebut. Sehingga, *sysadmin* tidak dapat mengidentifikasi serangan yang terjadi pada *website*. Oleh karena itu, diperlukan suatu sistem yang memanajemen dan memantau serangan melalui *log* yang tercatat oleh *web server* (Suharjo, 2015). Salah satunya dengan menggunakan metode SIM (*Security Information Management*).

SIM (*Security Information Management*) merupakan suatu sistem yang berfungsi untuk memantau semua *service* yang tercatat pada *log* di *server. Log* yang

banyak dan rumit dapat ditranslasikan menjadi sebuah informasi yang mudah dibaca, sehingga kegiatan *server* dapat terpantau oleh *sysadmin* (Pratama, Wijaya and D, 2016).

Oleh karena itu, pada penelitian ini dilakukan prediksi sebuah serangan yang ada pada *log* dengan menggunakan algoritma *Hidden Markov Model.* Dengan menggunakan *Hidden Markov Model,* dapat dilakukan *filtering* untung mengurangi data yang bukan serangan pada *log.* Algoritma ini menggunakan *state* yang tidak dapat diamati secara langsung (tersembunyi), tetapi hanya dapat diobservasi melalui suatu himpunan pengamatan lain dengan menggunakan perhitungan statistic (Cahyanto *et al.*, 2014).

1.1 Rumusan Masalah

Berdasarkan dari latar belakang yang sudah dijelaskan, maka dapat diangkat sebuah rumusan masalah yaitu.

1. Apakah *Hidden Markov Model* dapat memprediksi sebuah serangan atau tidak pada *log* Apache?
2. Berapa akurasi, presisi, dan *recall* yang dihasilkan pada *Hidden Markov Model* dalam memprediksi serangan?
3. Apakah *Security Information Management* dapat menampilkan informasi mengenai akses yang dilakukan *user* pada *website*?

1.2 Tujuan Penelitian

Tujuan dari penelitian ini ialah.

1. Untuk mengetahui informasi log secara *user-friendly* yang diperlihatkan dalam *dashboard* agar *sysadmin* dapat memantau kegiatan *website.*
2. Untuk melihat prediksi serangan atau tidak serangan yang terjadi pada *website* pada data *log* Apache.
3. Untuk melihat integrasi data pada *access log* dengan penambahan fitur pada *error log* dalam memprediksi serangan

1.3 Manfaat Penelitian

Manfaat dari penelitian ini ialah.

1. Untuk membantu *sysadmin* dalam memantau, memonitoring, dan menganalisis serangan yang terjadi pada *website.*

2. Untuk mengetahui hasil prediksi sistem dalam memprediksi sebuah serangan yang ada pada *log web server*.
3. Untuk melihat akurasi dari algoritma *Hidden Markov Model* dalam prediksi sebuah serangan atau tidak serangan.

1.4 Batasan Masalah

Batasan masalah yang ada pada penelitian ini yaitu.

1. Mendeteksi serangan hanya berfokus pada *log* dari *access log* dan *error log* dari *web server* Apache pada *website* SIAKAD UPNVJ dengan jangka waktu September 2019.
2. Program yang dibuat berupa *prototype*.
3. *Server* yang digunakan menggunakan OS Ubuntu *Server*.
4. Aplikasi belum bisa mengolah data secara *real-time.*
5. SIM hanya dapat memprediksi serangan atau tidak serangan, tidak dapat memprediksi jenis serangan yang terjadi.
6. Integrasi yang dilakukan hanya pada *access log* dan *error log* pada *log web server Apache web application*
7. SIM tidak dapat mendeteksi *IP Address* yang berasal dari VPN (*Virtual Private Network*) ataupun dari *Proxy Server*.

1.5 Luaran yang Diharapkan

Luaran yang diharapkan pada penelitian ini ialah hasil prediksi *Hidden Markov Model* melalui *access log* dan *error log* apakah terdapat sebuah serangan atau tidak dan akurasi dari *Hidden Markov Model* terhadap *access log* dan *error log* dan juga hasil tampilan *dashboard* dari *log web server*.

1.6 Sistematika Penulisan

Berikut merupakan sistematika penulisan berupa gambaran secara terperinci mengenai tiap bab pada penulisan yang menjelaskan kesinambungan tiap bab satu sama lain yang akan dijelaskan sebagai berikut :

BAB 1 : PENDAHULUAN

Pada Bab ini berisi Latar Belakang, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, Batasan Masalah, Luaran yang Diharapkan, dan Sistematika Penulisan.

BAB 2 : LANDASAN TEORI

Pada Bab II Landasan Teori berisi tentang teori-teori mendasar, referensi jurnal, dan metode yang digunakan dalam penelitian ini.

BAB 3 : METODOLOGI PENELITIAN

Pada Bab III Metodologi Penelitian berisi tentang kerangka pikir, alur metode dalam memproses penelitian ini, serta segala metode yang terdapat dalam penelitian ini.

BAB 4 : HASIL DAN PEMBAHASAN

Pada Bab IV Hasil dan Pembahasan berisi tentang penjelasan mengenai proses pengolahan data dan pembuatan model untuk sistem, lalu pembahasan tentang analisis hasil pengujian dari data yang sudah diolah pada penelitian ini.

BAB 5 : PENUTUP

Pada Bab V Penutup berisi tentang kesimpulan dari hasil dari penelitian yang dilakukan pada bab 4 (empat) dan juga saran yang dapat digunakan sebagai acuan agar sistem dapat diperbaharui lebih baik dan lebih dinamis.

DAFTAR PUSTAKA

RIWAYAT HIDUP

LAMPIRAN