SKRIPSI

DETEKSI SERANGAN *MALWARE* DENGAN *DOMAIN GENERATION ALGORITHM* BERDASARKAN ANALISIS *TRAFFIC* DNS

ANDI TANGGUH KIPPI NUSANTARA

NPM: 2015730017

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2020

UNDERGRADUATE THESIS

DOMAIN GENERATION ALGORITHM MALWARE ATTACK DETECTION BASED ON DNS TRAFFIC ANALYSIS

ANDI TANGGUH KIPPI NUSANTARA

NPM: 2015730017

DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2020

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

DETEKSI SERANGAN *MALWARE* DENGAN *DOMAIN GENERATION ALGORITHM* BERDASARKAN ANALISIS *TRAFFIC* DNS

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 17 Juni 2020

ANDI TANGGUH KIPPI NUSANTARA
NPM: 2015730017

LEMBAR PENGESAHAN

DETEKSI SERANGAN *MALWARE* DENGAN *DOMAIN GENERATION ALGORITHM* BERDASARKAN ANALISIS *TRAFFIC* DNS

ANDI TANGGUH KIPPI NUSANTARA

NPM: 2015730017

Bandung, 17 Juni 2020

Menyetujui,

Pembimbing

Chandra Wijaya, M.T.

Ketua Tim Penguji

Anggota Tim Penguji

Pascal Alfadian, Nugroho, M.Comp.

Lionov, Ph.D.

Mengetahui,

Ketua Program Studi

Mariskha Tri Adithia, P.D.Eng

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

DETEKSI SERANGAN *MALWARE* DENGAN *DOMAIN GENERATION ALGORITHM* BERDASARKAN ANALISIS *TRAFFIC* DNS

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 17 Juni 2020

ANDI TANGGUH KIPPI NUSANTARA
NPM: 2015730017

ABSTRAK

Teknik penanganan terhadap penyerangan *botnet* telah berkembang. Namun, penyerang menemukan teknik lain dalam menghindari pendeteksian dan blokade sehingga dapat bertahan hidup lebih lama. *Domain Generation Algorithm* (DGA) merupakan salah satu algoritma yang digunakan oleh penyerang untuk membangkitkan domain dengan karakteristik dan pola penamaan yang acak dalam jumlah besar, tetapi hanya sebagian dari domain tersebut yang digunakan *botnet* untuk berkomunikasi dalam kanal *command and control* (C&C). Penyerang memanfaatkan DGA untuk dapat berkomunikasi lebih lama dalam kanal C&C. Adanya pembangkitan domain dengan karakteristik dan pola penamaan yang acak dalam jumlah besar membuat penyerang dapat dengan mudah mengganti domain untuk komunikasi dengan domain lain jika pada suatu waktu terjadi pemblokiran domain yang sedang digunakan.

Teknik terbaru untuk mendeteksi serangan DGA adalah dengan menganalisis *response Non-eXistent Domain* (NXDOMAIN). NXDOMAIN merupakan salah satu jenis DNS *response* yang memiliki RCODE bernilai 3, yang menunjukkan domain tersebut tidak tersedia di DNS *server*. NXDOMAIN dapat dimanfaatkan dalam pendeteksian karena sebagian domain hasil *malware* dengan DGA yang tidak aktif akan menghasilkan *response* NXDOMAIN saat *botnet* ingin mencari domain yang menjadi titik temu komunikasi dalam C&C. Analisis dapat dilakukan dengan menangkap dan melakukan *parsing traffic* DNS *query* dan *response*, sehingga *response* NXDOMAIN yang dimiliki tiap domain dapat digunakan untuk menentukan apakah suatu domain merupakan hasil serangan DGA atau tidak.

Skripsi ini membuat sebuah perangkat lunak untuk mendeteksi domain serangan DGA dengan menganalisis jumlah NXDOMAIN yang diberikan tiap domain. Metode yang digunakan adalah dengan menghitung jumlah *response* NXDOMAIN dan jika suatu domain memiliki jumlah *response* NXDOMAIN lebih dari dua akan dianggap sebagai domain hasil *malware* dengan DGA. Hasil yang diberikan adalah sekumpulan domain dan rekapitulasi jumlah domain DGA dan non-DGA dari penangkapan *traffic* DNS menggunakan perangkat lunak yang dibuat dan *dataset malware traffic* yang disediakan oleh *Stratosphere IPS*. Tingkat akurasi dari perangkat lunak yang dibuat diukur dengan menghitung nilai *precision*, *recall* dan *F1 score* pada sampel *pcap CTU-Malware-Capture-Botnet-91*, *CTU-Malware-Capture-Botnet-221-1*, dan *CTU-Malware-Capture-Botnet-25-4*. Nilai terbaik dihasilkan dari analisis terhadap sampel *CTU-Malware-Capture-Botnet-25-4* dengan *precision* sebesar 0,999, *recall* sebesar 0,994, dan *F1 Score* sebesar 0,997 yang menunjukkan metode dapat digunakan dalam mendeteksi domain serangan *malware* dengan DGA.

Kata-kata kunci: *botnet*, *Command and Control* (C&C), DNS, *Domain Generation Algorithm* (DGA), *Non-eXistent Domain* (NXDOMAIN)

ABSTRACT

Detection techniques for botnet attacks have evolved. However, the attacker found other techniques to avoid detection and blockade so that they can survive longer. Domain Generating Algorithm (DGA) is one of the algorithms used by attackers to generate a large number of domains with randomly named characteristics and patterns, but only part of the domains used by botnets to communicate in command and control (C&C) channels. Attackers use DGA to be able to communicate longer in C&C channels. The existence of domain generation with such characteristics and randomly named patterns in large numbers enables an attacker can easily change the domain for communication with other domains if there occurs a blockade of the domain that is being used.

The latest technique for detecting DGA attacks is to analyze the Non-eXistent Domain (NXDOMAIN) response. NXDOMAIN is a DNS response type that has an RCODE value of 3, which indicates this domain is not available on the DNS server. NXDOMAIN can be used in detection because some of the generated DGA domains are inactive that will produce an NXDOMAIN response when botnets want to find a domain as a communication meeting point the C&C. Analysis can be done by capturing and parsing DNS requests and responses traffic so that the NXDOMAIN responses given by each domain can be used to determine whether each domain is the result of a DGA attack or not.

This thesis makes software to detect DGA attack domains by analyzing the number of NXDOMAIN given by each domain. The method used is to count the number of NXDOMAIN responses and considers the domain as DGA if the domain has more than two NXDOMAIN responses. The results provided are a collection of domains and a recapitulation of the number of DGA and non-DGA domains from DNS capture traffic using the created software and malware traffic datasets provided by Stratosphere IPS. The accuracy of the software is counted by calculating the value of precision, recall, and F1 scores on the pcap file samples of CTU-Malware-Capture-Botnet-91, CTU-Malware-Capture-Botnet-221-1, and CTU-Malware-Capture-Botnet - 25-4. The best value is generated from the analysis of the CTU-Malware-Capture-Botnet-25-4 sample with a precision of 0.999, a recall of 0.994, and an F1 score of 0.997 which means the method used to analyze can be used in detecting domains with malware with DGA

Keywords: botnet, Command and Control (C&C), DNS, Domain Generation Algorithm (DGA), Non-eXistent Domain (NXDOMAIN)

Untuk diri sendiri, keluarga, teman-teman seperjuangan, dan semua yang telah mendukung

KATA PENGANTAR

Puji Syukur penulis panjatkan kepada Tuhan Yang Maha Esa, karena dengan rahmat dan karunia-Nya, penulis dapat menyelesaikan skripsi yang berjudul "Deteksi Serangan *Malware* dengan *Domain Generation Algorithm* Berdasarkan Analisis *Traffic* DNS". Penulis berharap skripsi dan perangkat lunak yang dibangun dapat berguna bagi orang yang membutuhkan dan juga dapat membantu bagi orang yang akan melanjutkan penelitian ini untuk selanjutnya. Pada kesempatan kali ini penulis mengucapkan terima kasih kepada:

- Bapak Chandra Wijaya, S.T., M.T., yang telah membantu membimbing penulis dalam proses pembuatan skripsi beserta perangkat lunak.
- Kepala Lab FTIS beserta teman-teman Admin Lab FTIS yang telah membantu menyediakan komputer dan perangkat keras untuk pengembangan dan pengujian perangkat lunak.
- Keluarga, khususnya orang tua dan ketiga kakak yang telah menyemangati penulis secara langsung maupun tidak langsung, serta membuat suasana senyaman mungkin.
- Ayu, Reyner, Sandy, Enrico, Richard, dan teman-teman Informatika yang selalu saling mendukung dalam pengerjaan skripsi.
- Pihak-pihak lain yang tidak bisa disebutkan satu persatu.

Bandung, Juni 2020

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR KODE PROGRAM

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Domain Name System (DNS) adalah infrastruktur penting dalam internet yang bertugas untuk menerjemahkan *hostname* (domain) ke alamat *Internet Protocol* (IP). Sebagai contoh domain dari www.unpar.ac.id akan diterjemahkan menjadi alamat IP 43.252.138.173. DNS dapat ditemukan di berbagai macam aplikasi yang menggunakan internet, seperti pengaksesan *website*, *email*, dan aplikasi lainnya. Dalam DNS dikenal pesan dengan format tertentu yang disebut dengan DNS *packet*. DNS *packet* memiliki dua macam format agar DNS *client* dan *server* dapat berkomunikasi, yaitu DNS *query* dan DNS *response*. DNS *query* digunakan saat *client* ingin meminta informasi dari suatu nama domain ke pada DNS *server*, dan DNS *response* adalah pesan balikan yang diberikan oleh DNS *server* kepada *client*. Sebagai sebuah infrastruktur penting dalam internet, DNS tidak terlepas dari bahaya yang berada dalam layanan internet, seperti penyerangan terhadap pengguna dalam berbagai aspek, salah satunya adalah adanya perangkat lunak jahat yang dapat merusak komputer pengguna, atau biasa disebut sebagai *malware*.

Malware juga dapat mencuri informasi penting pengguna yang ada dalam komputernya, sebagian *malware* dapat menyebar melalui internet hingga membentuk satu jaringan antar komputer yang terinfeksi. Jaringan tersebut dinamakan *botnet*, di mana komputer atau *bot* yang terhubung dapat dikontrol melalui sebuah kanal *command and control* (C&C).

Layanan internet dimanfaatkan oleh *botnet* untuk saling berhubungan antar *bot* dan *botmaster* yang mengontrol para *bot* tersebut. Ada berbagai macam teknik yang digunakan *botnet* untuk menghindari pendeteksian di internet, dengan teknik terbarunya adalah memanfaatkan *Domain Generation Algorithm* (DGA). DGA digunakan *botnet* untuk membangkitkan sekumpulan nama domain dengan kombinasi karakter secara acak dalam jumlah yang besar, namun hanya beberapa domain aktif yang digunakan untuk berkomunikasi dengan *botmaster*. Teknik ini dapat menyulitkan pihak keamanan dalam melakukan pendeteksian dan pemblokiran, seperti pada salah satu teknik *reverse engineering*. *Reverse engineering* sering digunakan untuk melawan dan menutup akses penyerang ke sebuah sistem dengan cara memperkirakan calon kumpulan domain yang akan dibangkitkan, namun teknik ini membutuhkan sampel dari *malware* dengan DGA dan membutuhkan waktu yang lama untuk dianalisis. Dengan memanfaatkan DGA, penyerang dapat menambah waktu pendeteksian dan analisis akibat selalu berubahnya nama domain.

Salah satu teknik pendeteksian yang populer dan sering digunakan adalah dengan melakukan pengecekan kombinasi karakter *alphanumeric* pada nama domain, namun saat ini muncul teknik baru yang memanfaatkan informasi *Non-eXistent Domain* (NXDOMAIN). NXDOMAIN dapat dimanfaatkan untuk melakukan pendeteksian karena pola dan karakteristik *bot* yang melakukan *query* terhadap sekumpulan domain-domain yang dibangkitkan oleh *bot* tersebut, yang mana hanya sebagian domain yang aktif, namun sebagian lain tidak sehingga hasil *query* akan menghasilkan *response* berupa NXDOMAIN. Dengan menganalisa *bot* yang melakukan *query* dan menghasilkan *response* NXDOMAIN, dapat ditentukan apakah domain tersebut merupakan domain DGA hingga mengetahui secara spesifik komputer yang terinfeksi.

Pada skripsi ini dibuat perangkat lunak untuk melakukan penangkapan informasi *traffic* DNS terhadap sampel komputer yang terinfeksi *malware* dengan DGA dan sampel *network traffic* dalam bentuk *file pcap*, kemudian menganalisis informasi tersebut untuk mengetahui kumpulan domain yang merupakan hasil pembangkitan DGA berdasarkan hasil *response* berupa NXDOMAIN. Perangkat lunak ini dibuat dalam bahasa Python dengan memanfaatkan library *pyshark* untuk melakukan proses penangkapan *packet*, dan *python-flask* sebagai antarmuka dari program.

1.2 Rumusan Masalah

Pada skripsi ini terdapat rumusan masalah sebagai berikut:

1. Bagaimana cara menangkap dan mengekstrak informasi *packet* pada *traffic* DNS?
2. Bagaimana cara kerja serangan *malware Domain Generation Algorithm* (DGA)?
3. Bagaimana cara menganalisis suatu domain merupakan hasil serangan *malware Domain Generation Algorithm* (DGA)?

1.3 Tujuan

Tujuan dari skripsi ini adalah sebagai berikut:

1. Menangkap dan mengekstrak informasi *packet* pada *traffic* DNS.
2. Menganalisis cara kerja dan pola serangan *malware Domain Generation Algorithm* (DGA).
3. Mengimplementasikan cara melakukan analisis untuk mengetahui apakah suatu domain merupakan hasil serangan *malware Domain Generation Algorithm* (DGA) atau bukan.

1.4 Batasan Masalah

Batasan masalah dari skripsi ini adalah sebagai berikut:

1. Perangkat lunak hanya dapat mengidentifikasi apakah suatu *domain* merupakan hasil DGA, tidak mengidentifikasi jenis *malware* yang menghasilkan *domain* DGA tersebut.
2. *File PCAP* yang digunakan sebagai sampel untuk eksperimen didapat dari situs *Stratosphere IPS*[1]

1.5 Metodologi

Langkah-langkah pengerjaan dari skripsi ini adalah sebagai berikut:

1. Melakukan studi literatur mengenai *Domain Name System* (DNS).
2. Melakukan studi literatur mengenai *malware*, *botnet*, dan *command and control* (C&C).
3. Melakukan studi literatur mengenai *Domain Generation Algorithm* (DGA).
4. Melakukan studi literatur mengenai teknik menganalisis pola informasi serangan DGA.
5. Melakukan perancangan dan implementasi perangkat lunak untuk menangkap *traffic* DNS dan melakukan *parsing* file *pcap* menggunakan *library pyshark* dan menganalisis domain hasil DGA.

[1] https://www.stratosphereips.org/datasets-malware

6. Melakukan perancangan dan implementasi perangkat lunak dengan antarmuka untuk melakukan penangkapan *traffic* DNS, mengunggah file *pcap*, dan menampilkan hasil analisis domain.

7. Melakukan pengujian terhadap perangkat lunak yang dibuat.

8. Menulis dokumen skripsi.

1.6 Sistematika Pembahasan

Sistematika pembahasan dalam skripsi ini adalah sebagai berikut:

1. Bab I
   Bab ini berisi latar belakang, rumusan masalah, tujuan, metodologi, dan sistematika pembahasan yang berisi ringkasan dari tiap bab secara umum.

2. Bab II
   Bab ini berisi teori-teori dasar mengenai *Domain Name System* (DNS), *malware*, *botnet* dan *command and control* (C&C), *Domain Generation Algorithm* (DGA), *Pyshark*, *Flask* dan *Flask-SocketIO*.

3. Bab III
   Bab ini berisi analisis teori mengenai analisis riset sejenis dan terkait, analisis metode pendeteksian, analisis sistem yang akan dibangun, deskripsi perangkat lunak, percobaan *library Pyshark*, analisis *Entity Relationship Diagram* (ERD), analisis *Use Case Diagram*, analisis diagram konteks, dan analisis *Data Flow Diagram* (DFD).

4. Bab IV
   Bab ini berisi perancangan sistem mengenai diagram alir sistem, perancangan basis data, perancangan modul, perancangan metode analisis domain DGA, dan perancangan antarmuka.

5. Bab V
   Bab ini berisi implementasi dan pengujian dari perangkat lunak.

6. Bab VI
   Bab ini berisi kesimpulan dari eksperimen yang dilakukan dalam skripsi beserta saran terhadap pengembangan perangkat lunak.