SKRIPSI

PEMBANGUNAN PERANGKAT LUNAK PENDETEKSI MALWARE DENGAN PENDEKATAN GRAYSCALE IMAGE

Ivan Limosi

NPM: 6181801032

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2023

UNDERGRADUATE THESIS

MALWARE DETECTION SOFTWARE DEVELOPMENT WITH GRAYSCALE IMAGE APPROACH

Ivan Limosi

NPM: 6181801032

DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2023

LEMBAR PENGESAHAN

PEMBANGUNAN PERANGKAT LUNAK PENDETEKSI MALWARE DENGAN PENDEKATAN GRAYSCALE IMAGE

Ivan Limosi

NPM: 6181801032

Bandung, 06 Juli 2023

Menyetujui,

Pembimbing

Digitally signed by Chandra Wijaya

Chandra Wijaya, M.T.

Ketua Tim Penguji

Digitally signed by Elisati Hulu

Elisati Hulu, M.T.

Anggota Tim Penguji

Digitally signed by Lionov

Lionov, Ph.D.

Mengetahui,

Ketua Program Studi

Digitally signed by Mariskha Tri Adithia

Mariskha Tri Adithia, P.D.Eng

PERNYATAAN

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

PEMBANGUNAN PERANGKAT LUNAK PENDETEKSI MALWARE DENGAN PENDEKATAN GRAYSCALE IMAGE

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 06 Juli 2023

Ivan Limosi
NPM: 6181801032

ABSTRAK

Malware merupakan perangkat lunak yang bekerja dengan memasuki sebuah perangkat tanpa sepengetahuan pengguna dan akan menyebabkan kerusakan pada perangkat tersebut, untuk menjaga keamanan pada perangkat malware tersebut perlu untuk dideteksi terlebih dahulu sebelum dapat merusak perangkat. Pendeteksian malware merupakan aspek vital dalam keamanan komputer, karena membantu melindungi terhadap ancaman siber dan menjaga keamanan data sensitif. Metode tradisional untuk mendeteksi malware seringkali berpatokan pada identifikasi fitur spesifik didalam kode yang dieksekusi, namun metode-metode ini dapat dengan mudah dikelabui oleh para penyerang dengan memodifikasi kode untuk mengelakkan detesi.

Dalam penelitian ini, dilakukan sebuah pendekatan untuk mendeteksi malware menggunakan grafik entropi dan *grayscale image*. Pendekatan ini melibatkan konversi file menjadi *grayscale image* dan kemudian menganalisa *grayscale image* tersebut menggunakan grafik entropi. Proses konversi tersebut memungkinkan perangkat lunak untuk menangkap struktur dari keseluruhan kode, sementara analisis pada grafik entropi dapat digunakan untuk mengidentifikasi pola di dalam gambar yang mungkin merupakan indikasi dari sebuah malware. Perangkat lunak yang dihasilkan dapat mengkonversi file menjadi *grayscale image*, membuat grafik entropi dari *grayscale image* tersebut dan melihat kemiripan dari file yang diunggah dengan bank data yang ada. Evaluasi dari pendekatan ini pada beberapa dataset yang berisi file berbahaya dan juga tidak berbahaya mendapatkan bahwa perangkat lunak dapat memiliki nilai akurasi diatas 50% untuk beberapa dataset.

Hasil dari pengujian akhir menunjukan bahwa penggunaan grafik entropi dan *grayscale image* untuk mendeteksi malware cukup efektif dan dapat menjadi salah satu teknik yang berguna untuk pendeteksian malware kedepannya.

Kata-kata kunci: Malware, entropi, keamanan cyber, gambar skala abu-abu, ransomware

ABSTRACT

Malware is software that operates by infiltrating a device without the user's knowledge and can cause damage to the device. To ensure the security of a device, malware needs to be detected before it can harm the device. Malware detection is a critical aspect of computer security as it helps protect against cyber threats and safeguard sensitive data. Traditional methods for detecting malware often rely on identifying specific features within the executed code. However, these methods can easily be evaded by attackers who modify the code to avoid detection.

In this research, an approach is taken to detect malware using entropy graphs and grayscale images. This approach involves converting files into grayscale images and then analyzing these grayscale images using entropy graphs. The conversion process allows the software to capture the structure of the entire code, while the analysis of entropy graphs can be used to identify patterns in the images that may indicate the presence of malware. The resulting software can convert files into grayscale images, generate entropy graphs from these images, and compare the uploaded file's similarity with existing database files. Evaluations of this approach on various datasets containing both malicious and benign files show that the software can achieve accuracy above 50% for some datasets.

The final test results demonstrate that the use of entropy graphs and grayscale images for malware detection is quite effective and can be a useful technique for future malware detection.

Keywords: Malware, entropy, cyber security, grayscale image, ransomware

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena dengan rahmat dan karunia-Nya, penulis dapat menyelesaikan skripsi ini sebagai salah satu syarat kelulusan untuk memperoleh gelar sarjana. Pada pengerjeaan skripsi ini penulis menyadari bahwa masih terdapat kekurangan yang disebabkan oleh keterbatasan kemampuan dan pengetahuan yang dimiliki oleh penulis, namun dukungan dari keluarga serta teman-teman selalu ada menyertai hingga skripsi ini berhasil diselesaikan. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada :

1. Orang tua dan saudara penulis yang selama ini selalu memberikan dukungan selama penulis mengerjakan skripsi ini.
2. Bapak Chandra Wijaya selaku dosen pembimbing yang sudah menemani dan memberikan banyak bantuan berupa kritik, saran, dan nasihat dalam proses pembuatan skripsi ini.
3. Bapak Elisati Hulu dan Bapak Lionov selaku dosen penguji yang telah bersedia memberikan kritik dan saran untuk perbaikan dari skripsi ini.
4. Christopher William sebagai teman sesama pejuang skripsi yang memberikan bantuan moral dan juga menemani dalam proses pembuatan skripsi ini.
5. Rekan-rekan Teknik Informatika 2018 yang tidak dapat disebutkan satu-per-satu

Akhir kata penulis berterimakasih dan memohon maaf sebesar-besarnya jika masih terdapat banyak kesalahan dan juga kelalaian yang dilakukan oleh penulis dan juga kepada pihak-pihak yang sudah membantu penulis dalam penyelesaian skripsi ini tetapi tidak tertuliskan oleh penulis. Semoga penelitian ini dapat menjadi sebuah inspirasi bagi peneliti berikutnya maupun pengetahuan yang berguna bagi pembaca.

Bandung, Juli 2023

Penulis

DAFTAR ISI

DAFTAR GAMBAR

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Cybersecurity merupakan bidang yang kian penting seiring perkembangan teknologi informasi dan komunikasi. Sejalan dengan peningkatan ketergantungan manusia terhadap teknologi, berbagai ancaman dan risiko keamanan digital juga semakin tinggi. Salah satu ancaman tersebut adalah *malware* yang dirancang untuk menyusup atau merusak sistem komputer tanpa persetujuan dari pengguna.

Malicious Software atau yang biasa disebut *Malware* adalah sebuah perangkat lunak yang jika aktif akan merusak kinerja perangkat keras atau komputer kita seperti memberikan pekerjaan yang sangat berat kepada prosesor dari komputer tersebut yang mengakibatkan perangkat keras menjadi lebih lambat, mengambil data-data penting yang ada di komputer kita, mendapatkan akses ke komputer kita, dapat menampilkan iklan-iklan yang mengganggu di komputer kita, dapat memberatkan kinerja prosesor walaupun tidak banyak aplikasi yang terbuka dan yang paling umum terjadi adalah adanya degradasi performa pada komputer kita[1]. Pada umumnya *Malware* ini akan muncul dari unduhan pengguna dengan beberapa bentuk berkas yang memiliki berbagai jenis atau bisa juga berasal dari *Flashdisk* yang mengandung *file malware* yang kita masukkan ke dalam komputer kita.

Malware memiliki banyak jenis, berikut adalah beberapa contoh dari jenis-jenis *malware* yang sudah diketahui yaitu Virus, Worm, Trojan Horse, Spyware, Scareware, Adware, Botnet, dan Ransomware[2]. Setiap jenis dari *malware* tersebut memiliki perbedaan dalam karakteristik dan cara merusak komputer pengguna, misalnya pada umumnya *malware* bertipe virus akan merusak komputer dengan cara memasukkan kode dirinya kedalam program yang ada didalam komputer tersebut dan akan melakukan hal yang sama jika ada seseorang yang membuka program tersebut, namun untuk *malware* bertipe adware akan menampilkan iklan yang tidak diinginkan dalam komputer tapi tidak akan memasukkan kode dirinya kedalam program lain. Untuk mendapatkan data dari *file malware* untuk pengujian tersebut dibutuhkan yang namanya *malware repository.*

Malware repository adalah sebuah tempat penyimpanan dari *file malware* yang masih aktif dan berbahaya.Pada pengerjaan skripsi ini akan digunakan *repository malware* yang bernama theZoo. TheZoo adalah sebuah proyek yang ditujukan untuk membuat kesempatan untuk pengguna agar dapat membuat analisis tentang *malware* yang masih aktif dan dapat diakses oleh publik. TheZoo pada awalnya dibuat oleh Yuval tisf Nativ pada tanggal 9 januari 2014 dan saat ini sedang diurus oleh Shahak Shalev. Dikarenakan pengumpulan data *malware* dapat dibilang susah, theZoo menyediakan hampir seluruh versi dari *malware* jenis apapun yang masih ada dan sudah dikumpulkan sejak tahun 2014 dalam sebuah repository github agar dapat diakses dengan aman. Data malware yang digunakan untuk pembandingan akan disimpan dalam folder lokal pada komputer yang digunakan untuk pembangunan. Selain *file malware* ada juga *file non malware* atau *benign file* yang digunakan sebagai pembanding bersama dengan *file malware.*

Pada saat ini ada beberapa cara untuk mendeteksi keberadaan *file* tersebut pada komputer seperti *signature based detection*, *static file analysis* dan lain-lain. Cara yang digunakan dalam pembangunan perangkat lunak ini adalah dengan menggunakan pendekatan *Grayscale Image. Grayscale Image*

adalah sebuah gambar yang disetiap pixel nya akan merepresentasikan sebuah tingkat warna abu-abu yang dimulai dari putih sampai dengan hitam dimana angka 0 akan merepresentasikan warna hitam dan angka 255 merepresentasikan warna putih. Gambar berikut adalah sebuah contoh *Grayscale Image*.

Gambar 1.1: Contoh *grayscale image*.

Setiap *byte* dari sebuah *file malware* akan direpresentasikan dalam bentuk grayscale image dan gambar tersebut akan digunakan untuk perbandingan dengan gambar dari *file malware* yang terdapat pada bank data untuk dilakukan perhitungan nantinya.

Untuk melakukan pembandingan *Grayscale Image* tersebut dapat dilakukan dengan menggunakan teknik *entrophy graph similarity* selain teknik ini juga terdapat beberapa teknik lain seperti *structural similarity index*, *normalized cross correlation* dan masih banyak lagi. Teknik *entrophy graph similarity* adalah teknik yang menggunakan angka entropi sebagai variabel yang digunakan untuk menentukan apakah kedua buah gambar tersebut mirip atau sama. Jadi jika nilai pada setiap pixel pada gambar yang ada cenderung mirip atau sama maka nilai dari entropi akan tinggi, sebaliknya jika nilai pada setiap pixel pada gambar yang ada itu berbeda-beda atau cenderung tidak seragam maka nilai dari entropi akan rendah.

Pada pembuatan skripsi ini bahasa pemrograman yang digunakan untuk membangun aplikasi pendeteksi malware adalah python. Bahasa pemrograman python merupakan salah satu bahasa pemrograman yang paling populer dan umum untuk digunakan [3]. Penggunaan python sebagai bahasa pemrograman pada skripsi ini dikarenakan python lebih mudah untuk dipelajari, diaplikasikan dan lebih fleksibel dibandingkan bahasa pemrograman yang lain karena dapat dijalankan pada berbagai *Operating System* seperti Windows, Linux dan lainnya. Perangkat lunak yang dibangun akan dapat mengunggah file apapun dan mengubahnya menjadi sebuah grayscale image, lalu grayscale image tersebut akan dapat dihitung nilai entropinya tiap baris dan dijadikan sebuah entropy graph, dan yang terakhir perangkat lunak dapat membandingkan file yang diunggah dengan seluruh file yang terdapat pada bank data.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, rumusan masalah pada skripsi ini adalah sebagai berikut :

1. Bagaimana cara kerja pendekatan *Grayscale Image* untuk memvisualisasikan *malware* tersebut?
2. Bagaimana cara membuat aplikasi untuk mendeteksi *malware* menggunakan pendekatan *Grayscale Image*?
3. Bagaimana cara menghitung persentase kemiripan suatu *malware* dengan keluarga tertentu dengan menggunakan teknik *entrophy graph similarity*?

1.3 Tujuan

Berdasarkan rumusan masalah yang telah ada, berikut adalah tujuan dari pembuatan skripsi ini :

1. Mempelajari cara kerja pendekatan *Grayscale Image* untuk memvisualisasikan *malware* tersebut.
2. Membangun aplikasi untuk mendeteksi *malware* menggunakan pendekatan *Grayscale Image*.
3. Menampilkan persentase dari perbandingan kemiripan suatu *malware* menggunakan teknik *entrophy graph similarity*.

1.4 Batasan Masalah

Pada skripsi ini terdapat beberapa batasan masalah yaitu :

1. Data *malware* yang digunakan untuk pengujian hanya berasal dari theZoo.
2. Database untuk penyimpanan data *malware* hanya disimpan dalam hardisk lokal.
3. Pengujian hanya dilakukan berdasarkan struktur dari *grayscale image* dan *entropy graph* dari sebuah *file*, tidak sampai menjalankan kode program ataupun membuka *file* dari *malware* yang sudah disediakan.

1.5 Metodologi

Berikut adalah metodologi penelitian yang dilakukan pada pembuatan skripsi ini :

1. Studi literatur tentang *malware*.
2. Mempelajari lebih dalam tentang bahasa pemrograman python.
3. Mempelajari tentang nilai *entropy*, dan *entropy graph*.
4. Mempelajari tentang teknik perhitungan *similarity* untuk *entropy graph*.
5. Melakukan pengujian terhadap seluruh *dataset*.
6. Mempelajari cara untuk mengubah *file* menjadi sebuah *grayscale image*.
7. Mempelajari cara untuk membuat *entropy graph* dari sebuah *file grayscale image*.
8. Menganalisis kebutuhan perangkat lunak.
9. Melakukan implementasi pengubahan *file* menjadi sebuah *grayscale image*.
10. Melakukan implementasi pembuatan sebuah *entropy graph* dari *grayscale image*.
11. Melakukan perbaikan pada aplikasi jika ditemukan *bug* atau error saat setelah di analisis.
12. Menyelesaikan keseluruhan dokumen skripsi.

1.6 Sistematika Pembahasan

Sistematika pembahasan pada skripsi yang dibuat kali ini adalah sebagai berikut:

1. Pada Bab 1 akan dibahas mengenai latar belakang masalah, rumusan masalah, tujuan dari pembuatan skripsi, batasan yang dimiliki oleh skripsi ini, metodologi penelitian, dan sistematika pembahasan.
2. Pada Bab 2 akan dibahas mengenai keseluruhan landasan teori yang akan digunakan dalam pengembangan aplikasi Pendeteksian *Malware* Dengan Pendekatan *Grayscale Image*. Pada bab ini akan banyak berisi teori tentang *Grayscale image, entropy graph, malware* , teknik perhitungan *similarity*, dan bahasa pemrograman yang digunakan yaitu python beserta semua *library* atau *framework* yang digunakan dalam pembuatan kode skripsi.
3. Pada Bab 3 akan dibahas mengenai analisis yang diperlukan sebelum memulai pembuatan perangkat lunak. Bagian ini akan terdiri dari Analisis Masalah, Analisis Malware, Analisis Entropy dan Entropy Graph, Gambaran Umum Perangkat Lunak, Analisis Perangkat Lunak, dan Analisis Data Flow Diagram.

4. Pada Bab4 akan dibahas mengenai perancangan yang akan dilakukan untuk membangun perangkat lunak. Bagian ini akan terdiri dari Interaksi Manusia dengan Perangkat Lunak, Perancangan Antarmuka, Perancangan Algoritma dan Perancangan Perangkat Lunak.
5. Pada Bab 5 akan dibahas mengenai implementasi dari perangkat lunak sesuai dengan perancangan yang sudah dibuat dan juga hasil pengujian dari perangkat lunak yang sudah dibuat.
6. Pada Bab 6 akan dibahas tentang kesimpulan dari keseluruhan pengerjaan skripsi ini dan juga perangkat lunak yang sudah dibangun berdasarkan bab 4 dan bab 5. Selain kesimpulan bab ini juga akan mengandung saran dari penulis untuk pengembangan penelitian yang akan datang.