**KECERDASAN BUATAN**

**FINAL PROJECT**

**UJIAN AKHIR SEMESTER**



Nama : Adelya Natasya Ashilah Dwilieska

NIM : 09011281924055

Kelas : SK5B

Dosen Pengampuh : Prof. Dr. Ir. Siti Nurmaini, M.T. PhD

**Jurusan Sistem Komputer**

**Fakultas Ilmu Komputer**

**Universitas Sriwijaya**

**2021**

**Deteksi Malware Menggunakan Metode Machine Learning**

1. **Pendahuluan**

Komputer forensik adalah investigasi dan teknik analisis komputer yang melibatkan tahapan identifikasi, persiapan, ekstraksi, dokumentasi dan interpretasi dari data yang terdapat pada komputer yang berguna sebagai bukti dari peristiwa cyber crime. Komputer forensik pada awalnya dilakukan dengan cara menganalisis media penyimpanan dari sebuah sistem yang dicurigai telah terlibat dalam sebuah tindak kejahatan, dimana biasanya sistem perlu dinonaktifkan kemudian dibuat imagekloning dari media penyimpanan sistem tersebut. Imageinilah yang dianalisis yang dapat digunakan sebagai barang bukti untuk keperluan investigasi lebih lanjut.

Data volatilekhususnya pada memori fisik atau RAM sebuah sistemmenggambarkan seluruh kegiatan yang sedang terjadi pada sistem tersebut. seperti Volatilitas memungkinkan penyelidik forensik mengidentifikasi dan menghubungkan berbagai komponen untuk menyimpulkan apakah kejahatan cyberitu dilakukan menggunakan malware atau tidak. Sebagian Alat Forensik yang berfungsi mendeteksi malwaresecara otomatis, tetapi harus selalu terhubung dengan internet, dan deteksi malwareyang dilakukan terbatas.

Untuk menerapkan algoritma machine learningdan otomatisasi langkah – langkah dasar. Keuntungan terbesar dari alat ini adalah, Pengguna dapat mendeteksi semua proses yang berjalan pada memory volatiledan tidak harus terkoneksi dengan internet dan Pengguna tidak perlu mengingat perintah, sintaknya atau bahkan ketika mau menggunakan perintah mana. Ini sangat berguna bagi mereka yang tidak lebih suka bekerja pada utilitas baris perintah karena mereka menghindari mengingat perintah. Solusi yang diusulkan disebut Memory VolatileForensik untuk deteksi malwaremenggunakan algoritma machine learning, adalah solusi untuk pengguna yang ramah dan akurat untuk mengatasi masalah diatas, juga menganalisis dan memberikan laporan akhir yang akurat.

1. **Landasan Teori**
2. Digital Forensik

Digital Forensik merupakan bagian dari ilmu forensik yang melingkupi penemuan dan investigasi materi (data) yang ditemukan pada perangkat digital.

1. Forensik Jaringan

Forensik Jaringan Merupakan ilmu keamanan komputer berkaitan dengan investigasi untuk menemukan sumber serangan pada jaringan berdasarkan bukti log, mengidentifikasi, menganalisis, serta merekonstruksi ulang kejadian tersebut.

1. Cybercrime

Cyber Crimeadalah segala macam penggunaan jaringan computer untuk tujuan kriminal dan atau kriminal berteknologi tinggi dengan menyalahgunakan kemudahan teknologi digital.

1. Memory Volatile

Memory Volatile adalah penyimpanan komputer yang hanya menyimpan datanya saat perangkat diaktifkan. Sebagian besar RAM (*Random Access Memory*) yang digunakan untuk penyimpanan primer di komputer adalah memori yang mudah menguap. RAM jauh lebih cepat untuk dibaca dan ditulis dibandingkan dengan jenis penyimpanan lain di komputer, seperti hard disk atau media yang dapat dipindahkan. Namun, data dalam RAM tetap ada ketika komputer sedang berjalan, akan tetapi sebaliknya ketika komputer dimatikan, RAM kehilangan datanya.

1. Machine Learning

Machine Learning merupakan kemampuan komputer untuk melakukan pembelajaran tanpa harus menjelaskan (programmed) secara eksplisit kepada komputer. Cara kerja dari machine learning sebenarnya berbeda-beda sesuai dengan teknik atau metode pembelajaran seperti apa yang kamu gunakan pada machine learning. Namun pada dasarnya prinsip cara kerja pembelajaran mesin masih sama, yaitu pengumpulan data, eksplorasi data, pemilihan model atau teknik, memberikan pelatihan terhadap model yang dipilih dan mengevaluasi hasil dari machine learning.

1. **Alat Yang Diusulkan**

Penelitian ini dilakukan untuk membantu Digital Forensic Investigator, diasumsikan yang tidak ahli dalam menganalis malware tetapi diperlukan untuk memiliki beberapa mekanisme yang dapat dengan mudah mereka identifikasi kehadiran malware apa pun dalam dump memory RAM. Alat yang diusulkan yaitu Memory Volatile Forensic menggunakan Algoritma Machine Learning.

Proses otomatis pertama yang dilakukan oleh alat ini adalah mengekstrak file dump dan menempatkan file dump yang valid di tempat yang tepat untuk dieksekusi lebih lanjut. Langkah selanjutnya adalah memilih profil sistem operasi yang disediakan oleh alat ini. Setelah memberikan informasi file image dan profil, alat ini mulai menjelajahi dan menganalisis dump proses pada memory volatile, proses ini berperan penting dalam mengidentifikasi serangan malware. Sebagian besar malware termasuk ransomware berbasis jaringan dan berfungsi sebagai botnet. Sebagian besar malware ini harus terhubung ke pusat kontrol untuk mengeksekusi perintah selanjutnya atau mengirim informasi penting atau rahasia.

Alat ini menggunakan algoritma Machine Learning sebagai proses pendeteksian malware untuk dapat mendeteksi malware secara offline dan tidak terbatas selama proses berjalan. Alat ini adalah wadah untuk menutupi dan mengotomatiskan semua langkah yang diperlukan oleh proses memori forensik dalam membantu digital forensic investigator. Pengguna akan mendapatkan hasil yang akurat tanpa mengetahui baris perintah yang diberikan tools volatility. Alat ini tersedia hanya untuk Windows dan mendukung file image dump memori dari Windows, Linux dan Mac.

1. **Metodologi**
2. Tahapan Penelitian

Merupakan langkah-langkah dalam melakukan penelitian, berikut tahapan penelitian yang dilakukan :

1. *Pengumpulan data :* Pengumpulan informasi dari sumber yang berkaitan dengan penelitian, studi literatur yaitu sumber-sumber dari jurnal, buku, internet, artikel dan lain-lain.
2. *Pengumpulan Alat dan Bahan :* Pengumpulan kebutuhan-kebutuhan yang digunakan pada penelitian, baik berupa perangkat keras, dan perangkat lunak yang mendukung dalam pembuatan alat forensik ini.
3. *Perancangan :* Membuat rancangan dan interface untuk menghubungkan framework volatilitas dengan algoritma machine learning.
4. *Pengujian :* Pengujian hasil proses deteksi malware menggunakan algoritma machine learning.
5. *Analisis :* Untuk menganalisis terhadap proses yang terinfeksi malware pada memory volatile.
6. Perancangan
7. Proses pemodelan algoritma klasifikasi

Algoritma machine learning digunakan untuk mengekstrak model yang didapat dari pengolahan data mentah (raw data) yang dapat digunakan untuk berbagai tugas dan tujuan. Cara kerja machine learning sama halnya dengan cara manusia belajar. Agar mesin, dalam konteks ini adalah komputer memiliki kemampuan belajar yang sama dengan manusia, maka perlu adanya proses klasifikasi sebelum menganalisis, menilai dan mengambil tindakan.

1. Pengambilan sampel DataSet

DataSet yang dikumpulkan yaitu sebanyak 41.323 file sampel dari hasil ekstraksi folder system32 di windows 7,8.1 dan 10, file-file dalam folder system32 diekstrak setelah instalasi OS dengan update terbaru. Sampel malware yang dapat dieksekusi diunduh dari situs. Jumlah total sampel malware adalah 96.724 yang mengandung PE-header. “pefile” yang merupakan salah satu dari modul python dipilih untuk mendeteksi PE-header pada suatu file dan mengekstraksi informasi header-PE dari file PE tersebut. Sehingga total jumlah didalam sampel yaitu 138.047.

Tabel berikut adalah atribut-atribut atau parameter yang berhubungan dengan dataset sebanyak 57 atribut.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ATRIBUT / *PARAMETER* YANG DIGUNAKAN** | | | | |
| Name | AddressOf  EntryPoint | MajorSub  systemVe rsion | SizeOfHe  apCommi t | Resources  MaxEntro py |
| md5 | BaseOfCod  e | MinorSub  systemVe rsion | LoaderFla  gs | Resources  MeanSize |
| Machine | BaseOfDat  a | SizeOfIm  age | NumberO  fRvaAnd Sizes | Resources  MinSize |
| SizeOfOp  tionalHea der | ImageBase | SizeOfHe  aders | SectionsN  b | Resources  MaxSize |
| Character istics | SectionAli gnment | CheckSu m | Sections MeanEntr  opy | LoadConf iguration  Size |
| MajorLin kerVersio n | FileAlignm ent | Subsyste m | Sections MinEntro py | VersionIn formation Size |
| MinorLin kerVersio n | MajorOper atingSyste mVersion | DllCharac teristics | Sections MaxEntro py | legitimate |
| SizeOfCo de | MinorOper atingSyste mVersion | SizeOfSta ckReserv e | Sections MeanRaw size | Resources MaxEntro py |
| SizeOfIni tializedDa ta | MajorImag eVersion | SizeOfSta ckCommi t | Sections MinRaws ize | Resources MeanSize |

Atribut atau parameter tersebut menunjukkan bahwa setiap file Portable Executable (PE) yang diekstrak akan diklasifikasikan dengan 57 atribut yang sebelumnya sudah dibuat.

1. **Hasil dan Pembahasan**

Dari percobaan yang telah dibuat dengan menggunakan algoritma machine learning dalam mengklasifikasi malware didalam dataset, hasil dalam pengetesan data training nantinya akan digunakan untuk melatih algoritma.

Pengujian dengan berbagai macam sampel profil sistem operasi

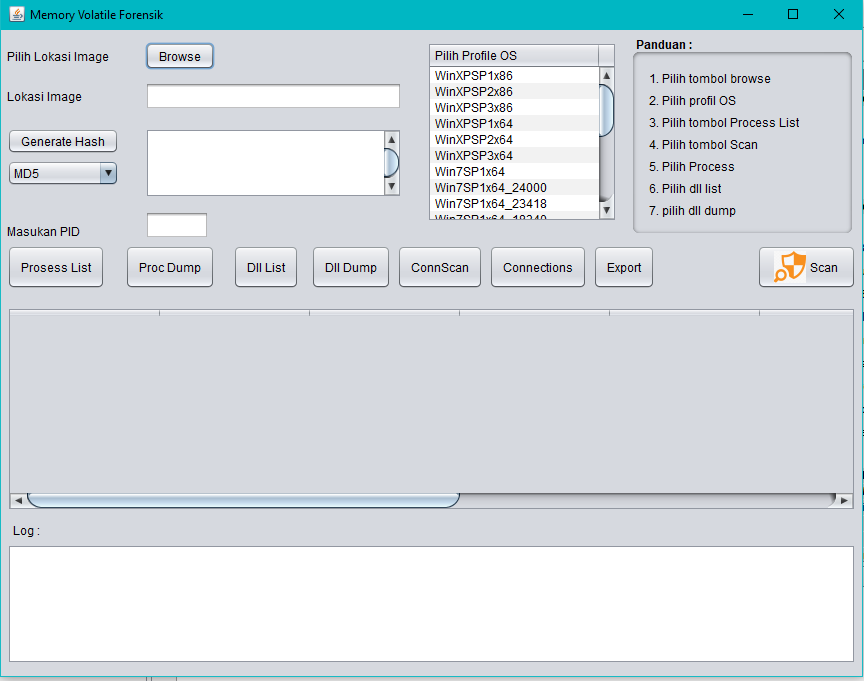
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **FEATU RE** | **PROFILE SISTEM OPERASI** | | | | | |
| Win XPS  P1x 86 | Win XPS  P2x 86 | Win 7SP  1x8 6 | Win 7SP  1x8 6 | Win 10x 64\_  149  39 | Win10 x86\_1 4939 |
| PROCE SS LIST | v | v | v | v | v | v |
| DLL | v | v | v | v | v | v |
| LIST |  |  |  |  |  |  |
| CONN SCAN | v | v | x | x | x | x |
| CONNE  CTION S | v | v | x | x | x | x |
| PROC DUMP | v | v | v | v | v | v |
| SCAN MALW ARE | v | v | v | v | v | v |

Ket : v = berjalan x = tidak berjalan

Setelah mendapatkan file, kemudian kita dapat menginstal alat ini setelah terinstal software pendukung yaitu :

* 1. JDK & JRE versi 1.8.0
  2. Python 2.7
  3. Pandas
  4. Numpy
  5. Pickle
  6. Scipy
  7. Scikit-learn
  8. Pefile

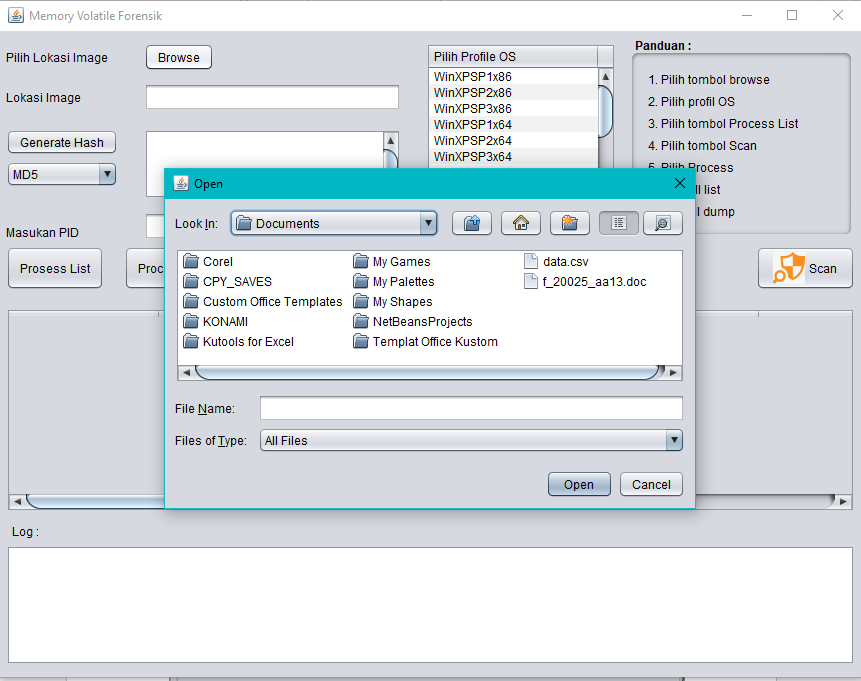
Setelah terinstal semua, alat ini bisa dijalankan.

Tampilan detail dengan fitur-fitur penting

Gambar diatas memiliki fitur-fitur penting yaitu :

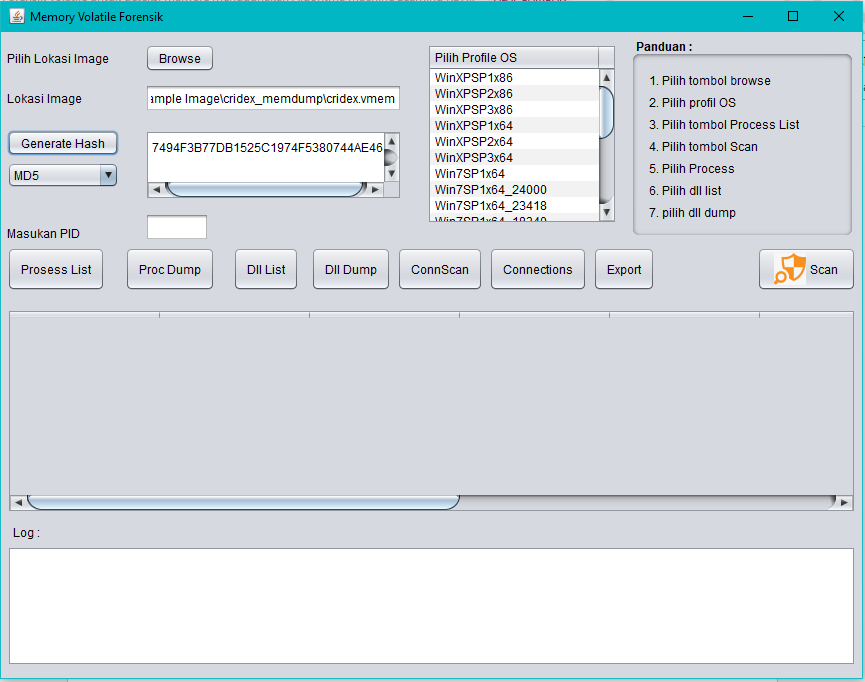
1. Browse yaitu untuk mengambil image yang akan dianalasis.
2. Profil OS yaitu jenis profil yang terdapat pada file image.
3. Generate Hash yaitu untuk mengetahui jenis dan nilai hash pada file image yang dimasukkan.
4. Process list yaitu untuk mengidentifikasi process yang berjalan.
5. Dll list yaitu untuk mengidentifikasi letak dll yang terhubung dengan process.
6. Connscan yaitu mengetahui remote IP yang digunakan.
7. Cconnections yaitu untuk mengetahui jenis nomor IP dan port yang digunakan.
8. Proc Dump yaitu untuk mengambil salah satu process / dll pada memory dan dapat diletakan pada local drive.
9. Scan yaitu untuk mengetahui process yang terinfeksi malware.

Memasukkan file dump RAM



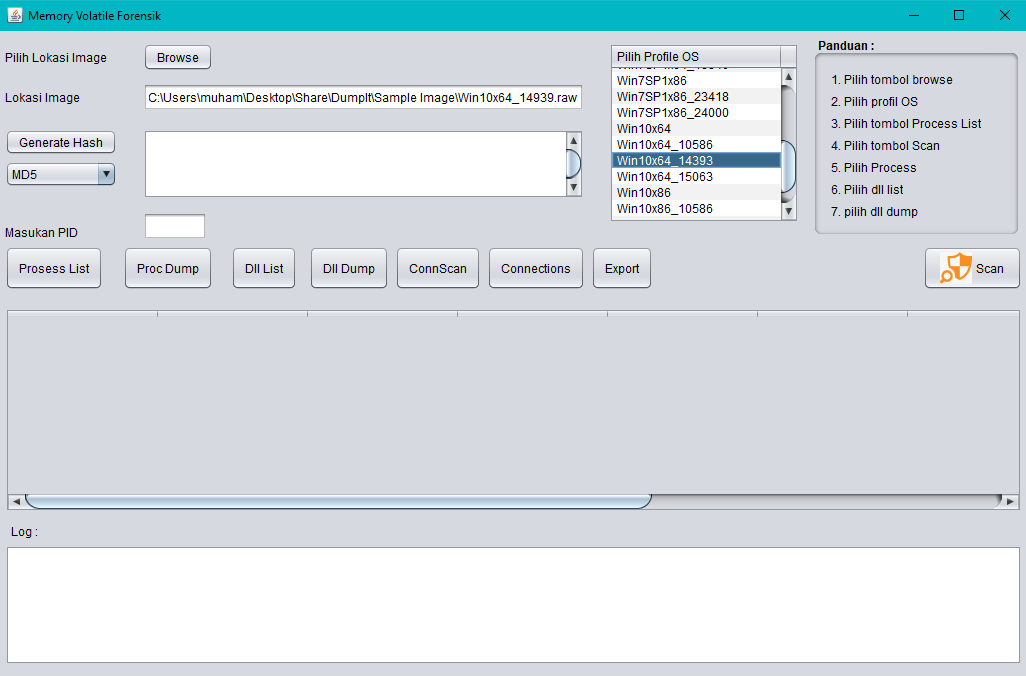
Pada gambar tersebut, pengguna dapat mengunggah file image dump memory menggunakan tombol browse. Karena ukuran dump memory mungkin dalam GB atau MB sesuai ukuran memory fisik yang diakusisi, pengguna diminta untuk mengklik tombol Analyze Dump, Alat ini akan mulai menganalisis file dump tersebut dan akan memperbarui menampilkan hasil dari forensik memori berupa profil proses yang sedang berjalan.

Menampilkan nilai Hash pada file image yang diinputkan



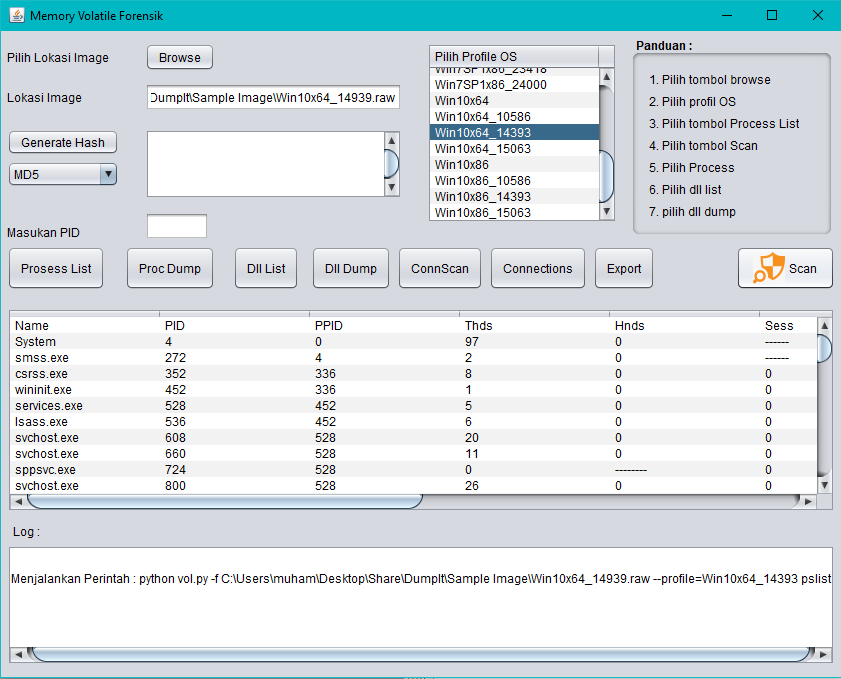
Setelah memasukan file image pengguna juga dapat melihat nilai hash suatu file, dengan cara memilih tombol generate hash. Nilai hash ini berfungsi sebagai keaslian suatu data, jika file image tersebut sudah ada yang memodifikasi, dapat diketahui pada bagian hash nya akan berubah, tidak akan sama dengan yang original.

Memilih Profil Sistem Operasi



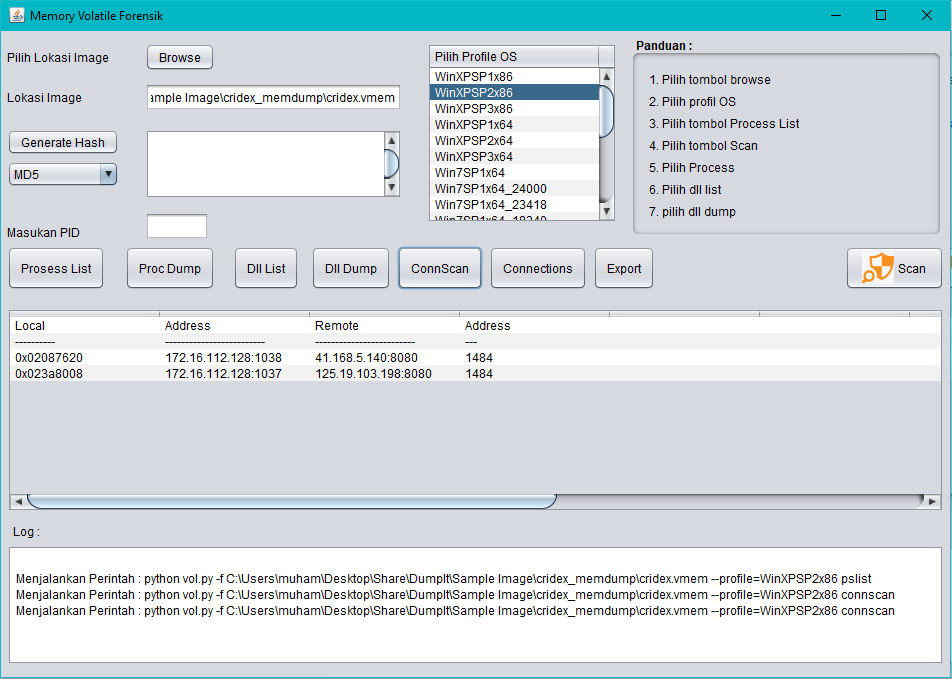
Pada gambar tersebut, pengguna dapat memilih profil sistem operasi sesuai dengan image yang diinputkan yang berfungsi menampilkan sistem operasi yang digunakan oleh file image tersebut, dari mulai sistem operasi windows, linux dan MacOS.

Informasi dan Proses Keseluruhan Dari Memory Volatile



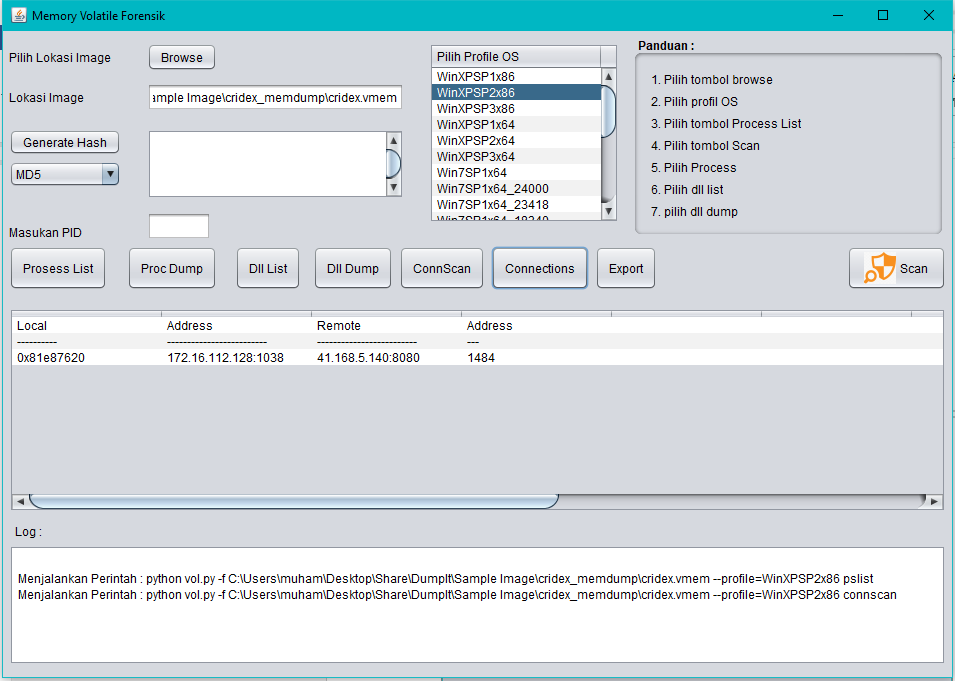
setelah analisis berhasil alat ini mengalihkan pengguna ke tombol yang lain di *interface* awal untuk menampilkan daftar proses dengan detail yang relevan seperti *thread, handle*, dan hal lain terkait dengan setiap proses.

Tampilan ConnScan pada jaringan



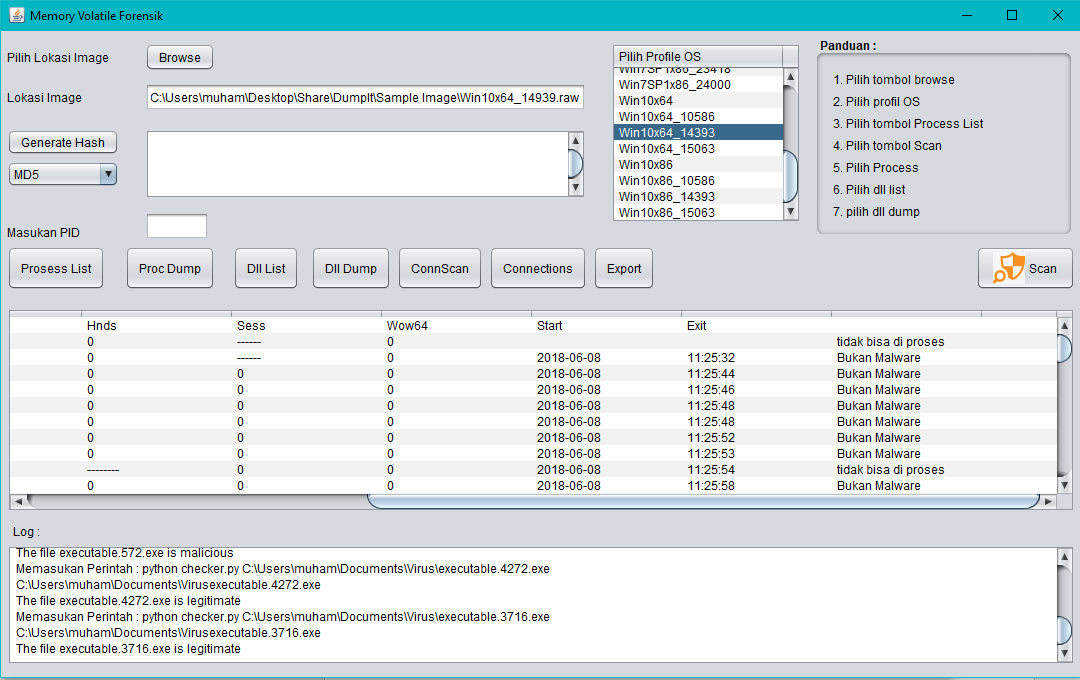
Fungsi Connscan yaitu *Scanning Memory* untuk mendapatkan koneksi *TCP*, termasuk *port* yang tertutup atau tidak terhubung *(Unlinked)* Prosess ini hanya berlaku pada file Windows XP atau 2003. Untuk indentifikasi remote ip tentu kita harus mengecek koneksi layaknya kita menggunakan Netstat, plugin yang digunakan untuk proses ini, kita bisa gunakan connections untuk melihat koneksi yang terbuka serta connscan untuk melihat koneksi TCP.

Tampilan Convection pada jaringan



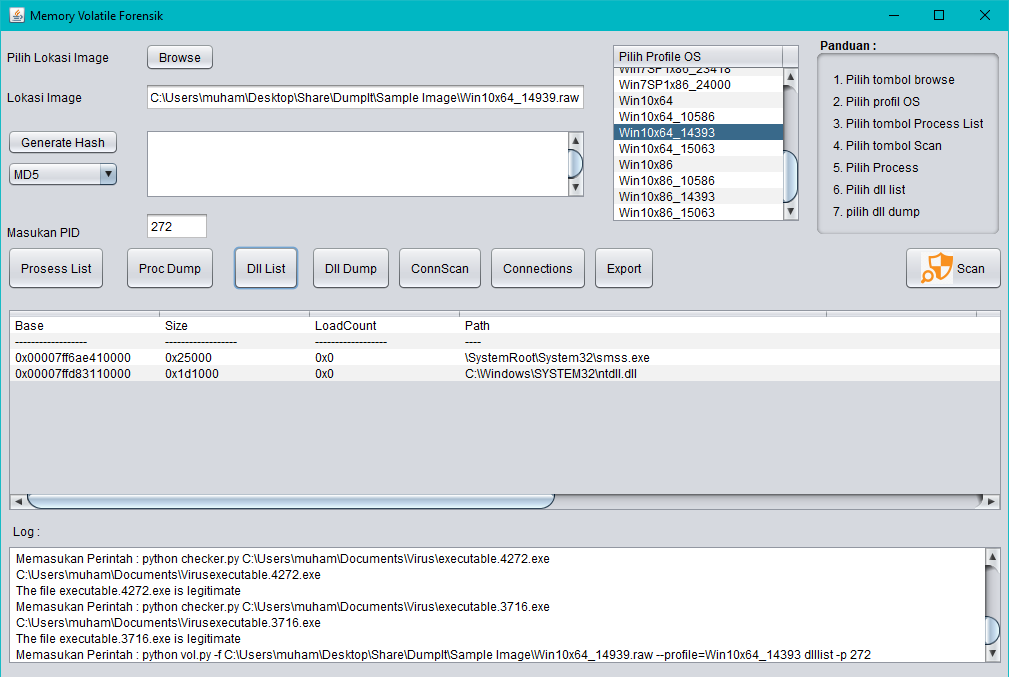
pengguna dapat mengetahui IP dan port yang terhubung pada komputer. Sehingga dapat diketahui port mana yang digunakan oleh *malware* untuk menyerang sistem, akan tetapi *tools* ini dapat menggunakan menu connection dan connscan jikalau tipe profilnya menggunakan WinXPSP1x86/64 dan WinXPSP2x86/64.

Proses Scan File Exacutable Yang Dicurigai Malware



Perspektif dalam menganalisis *malware* ini adalah proses yang mencurigakan dapat dianalisis dengan mengklik tombol scan. Itu akan memindai proses individu untuk virus, worm, Trojan. Ketika Anda mengklik tombol scan, itu akan menjalankan algorima *machine learning* untuk mendeteksi *malware* pada proses tersebut.

Menampilkan Dll List Dari Proses Yang Dipilih



pada proses ini pengguna dapat mengetahui file-file terinfeksi yang digunakan oleh *malware* dalam menyerang sistem operasi, dan dapat diekstrak melalui tombol *procdump*.

1. **Kesimpulan**

Hasil dari percobaan ini dapat disimpulkan bahwa algoritma machine learning sangat mempermudah dalam pengklasifikasian *malware* menggunakan dataset. Dan akan memberikan manfaat dan bantuan untuk penyelidik forensik dalam menganalisis memori *volatile* dan mendeteksi *malware* secara offline (tidak terhubung ke internet) yang mungkin ada pada *memory volatile*. Alat ini telah diuji dengan berbagai sampel *image* dan memberikan hasil yang akurat menurut dataset yang digunakan yaitu berisi file system32 dan *malware*. Akurasi dan keramahan pengguna alat ini akan membantu penyidik *forensic investigator* dan alat ini juga dapat mengurangi biaya pelatihan penyelidik forensik untuk menganalisis *malware*.

Padahal, alat ini memenuhi semua persyaratan dalam menganalisis *malware* dengan algoritma klasifikasi, masih ada ruang lingkup untuk pengembangan. Alat ini dapat diperluas dalam menganalisis *malware* secara detail. Alat yang dikerjakan saat ini hanya mendukung proses yang mengandung / terinfeksi *malware*, sehingga tidak dapat memberikan informasi yang lebih detail dari *malware* tersebut. Terakhir, dibuatkan sebuah GUI untuk proses otomatis ini dapat di operasikan pada sistem operasi yang lain seperti Linux dan Mac untuk lebih bervariasi.

***Referensi :***

1. *Jurnal Teknik Informatika dan Sistem Informasi (Fikri Bahtiar, Nur Widiyasono, Aldy Putra Aldya)*

[*https://www.academia.edu/43322455/Memory\_Volatile\_Forensik\_Untuk\_Deteksi\_Malware\_Menggunakan\_Algoritma\_Machine\_Learning*](https://www.academia.edu/43322455/Memory_Volatile_Forensik_Untuk_Deteksi_Malware_Menggunakan_Algoritma_Machine_Learning)

1. *Github*

[*https://github.com/surajr/Machine-Learning-approach-for-Malware-Detection*](https://github.com/surajr/Machine-Learning-approach-for-Malware-Detection)