Hafta 11 - CNN Zararlı Yazılım Analizi/ LSTM Parola Oluşturma

BGM 565 - Siber Güvenlik için Makine Öğrenme Yöntemleri

Bilgi Güvenliği Mühendisliği
Yüksek Lisans Programı

Dr. Ferhat Özgür Çatak
ozgur.catak@tubitak.gov.tr

İstanbul Şehir Üniversitesi
2018 - Bahar

İçindekiler

İçindekiler

LSTM Parola Oluşturma I

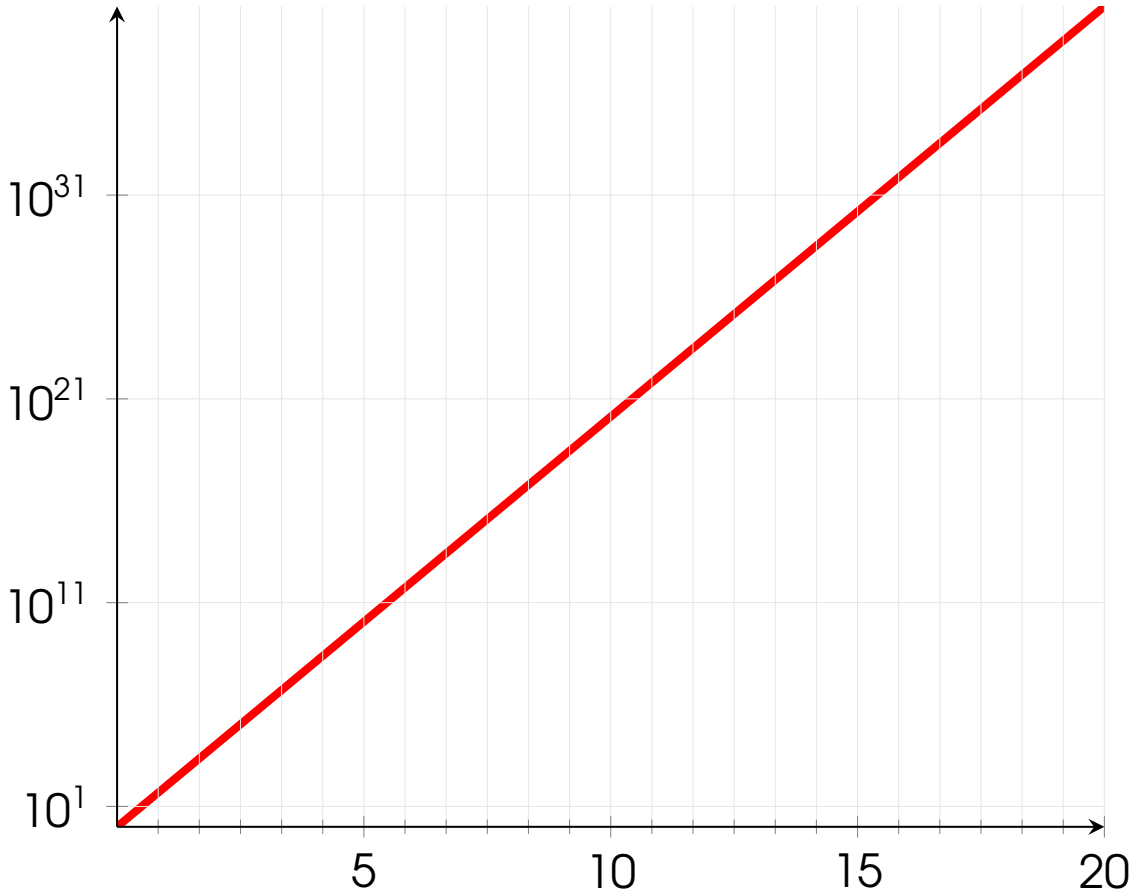
Parola

- Kimlik doğrulama sistemlerinde en çok kullanılan bileşen.
- Kimlik doğrulamada, parola politkası tanımlansa bile kullanıcılar *en kolay hatırlanacak* parolayı seçmeye eğilimlidirler.
- **qwertyasdfgzxcvb** şeklinde 16 karakterden oluşan bir parola kaba-kuvvet saldırıları için oldukça zor olmasına rağmen, belirli bir düzen (klavye üzerinde yer alan ilk 3 sıra karakterler) olması sebebiyle bir desene sahiptir.

Neden İhtiyacımız Var?

- Elde edilen bir hashin çözümlenmesi
- **John the ripper ???**

LSTM Parola Oluşturma II



Şekil: 103 karakter sözlüğe sahip parola boyutuna göre arama uzayının değişimi

LSTM Parola Oluşturma III

Table: Kullanıcı parolalarının bileşenleri

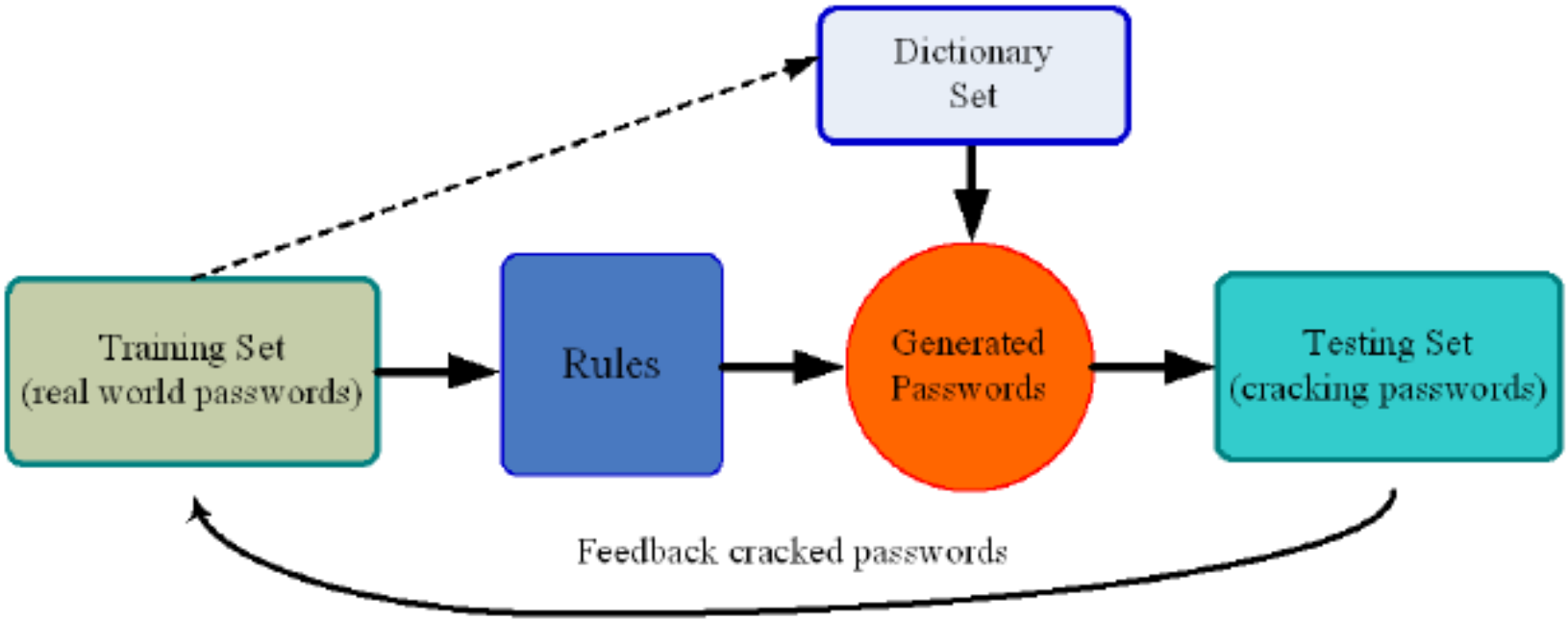
Tip	Toplam	Bileşenler
Rakam	10	0123456789
Büyük harf	31	ABCÇDEFGĞHIİJKLMNOÖPQRSŞTUVWXYZ
Küçük harf	31	abcçdefgğhiijklmnoöpqrsştuvwxyz
Özel	31	~'!@#$%^\&*()_-=[]|;':',./<>?

Tahmin - Üretken Modeller
Predictive - Generative Modeling

Üretken Modeller

- **Üretken model**: denetimsiz öğrenmeyi kullanarak her türlü veri dağıtımını öğrenmenin bir yoludur.
- Bazı varyasyonlarla yeni veri noktaları oluşturmak için eğitim setinin gerçek veri dağılımını öğrenmeyi amaçlamaktadır.
- **LSTM ağları**: text generation character-by-character
  - Karakterler arasındaki bağımlılıkları ve dizilerdeki karakterlerin koşullu olasılıklarını öğrenilmesi
  - Böylece tamamen yeni ve orijinal karakter dizileri oluşturulması.

Model I


Dictionary Set
Training Set (real world passwords)
Rules
Generated Passwords
Testing Set (cracking passwords)
Feedback cracked passwords

Model II

- **Rockyou veri kümesi**: 14.344.391 parola

```
168998348              umpalumpa1    alianzateamo
lingy7                 m1a2sep       samsam15
jessica29              raggi1983     jordhud
kepi15                 pajaroloco1   9841321
hotnica1 0ee2s28E35     SHY1
107Sherdale            blah1         84136
trinitynathaniel        pinguinitoz
```

Model III

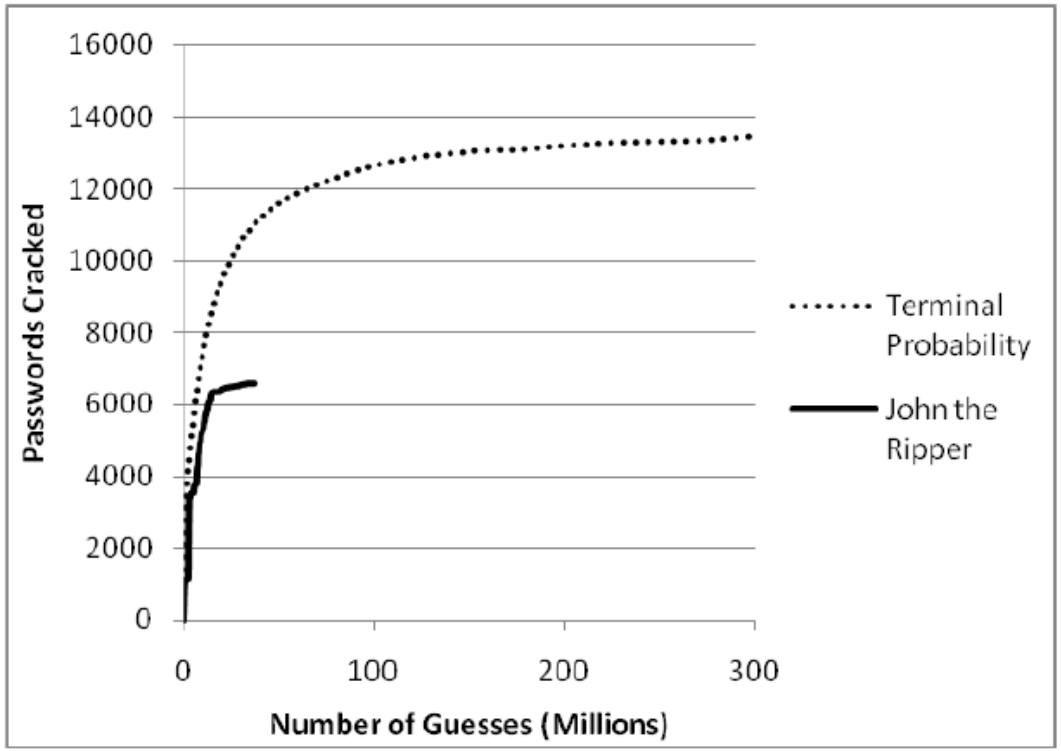
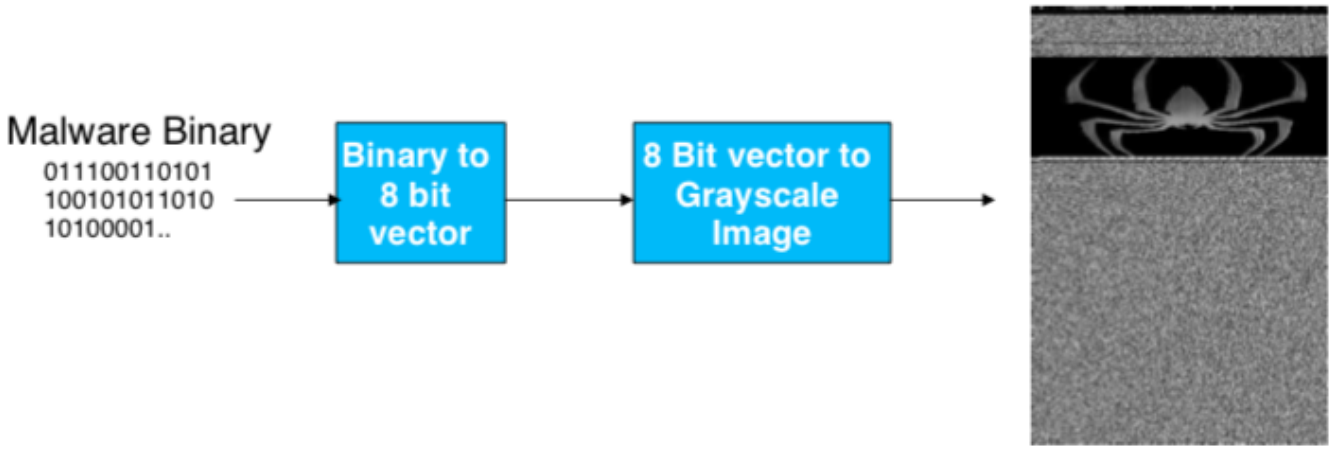


Fig. 4.4.6. Number of Passwords Cracked Over Time. Trained on the MySpace Training List. Tested on the MySpace Test List

İçindekiler

CNN Zararlı Yazılım Tespiti I

Alternatif zararlı yazılım analiz yöntemi



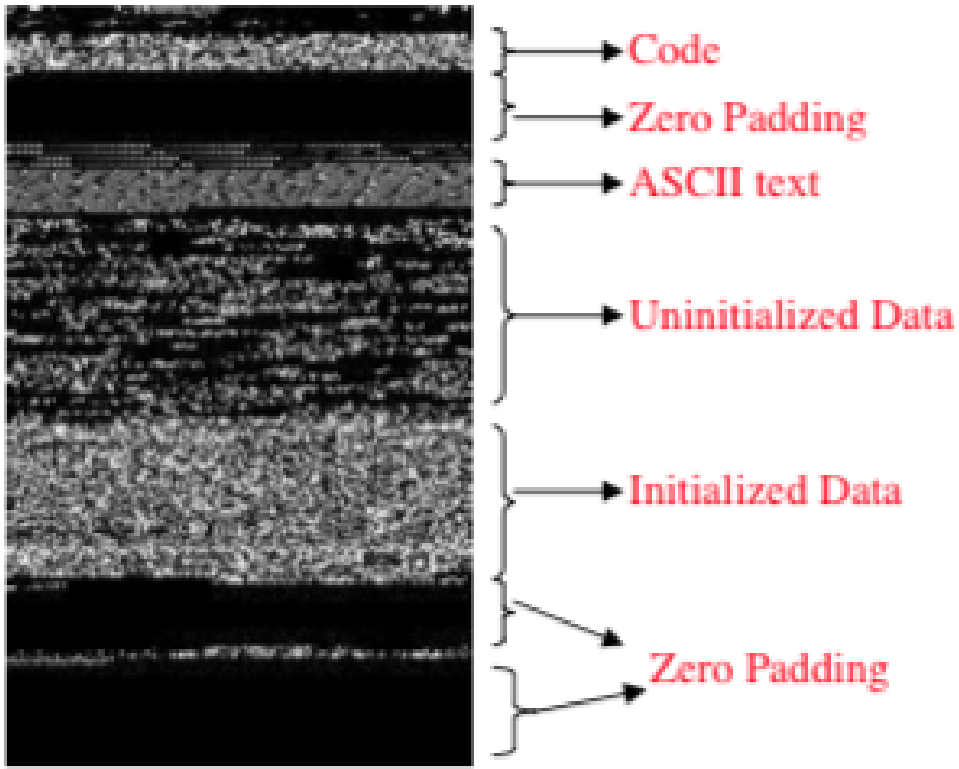
Malware Images

- **Binary gösterim**: Farklı bölümler imge olarak görüntülebilinir.
- Zararlı yazılım geliştiricileri, variant oluşturmak için orjinal kodu bazı bölümlerini değiştirmektedir.
- Imge kullanarak bu ufak değişimler kolaylıkla fark edilebilir.
- Zararlı yazılımlar benzer ailelere ait olanlar benzer resimler oluşturacaklardır.

CNN Zararlı Yazılım Tespiti II

Alternatif zararlı yazılım analiz yöntemi

Information that we can obtain from images



CNN Zararlı Yazılım Tespiti III

Alternatif zararlı yazılım analiz yöntemi

File Size Range	Image Width
<10 kB	32
10 kB – 30 kB	64
30 kB – 60 kB	128
60 kB – 100 kB	256
100 kB – 200 kB	384
200 kB – 500 kB	512
500 kB – 1000 kB	768
>1000 kB	1024

Packed and Obfuscated Malware

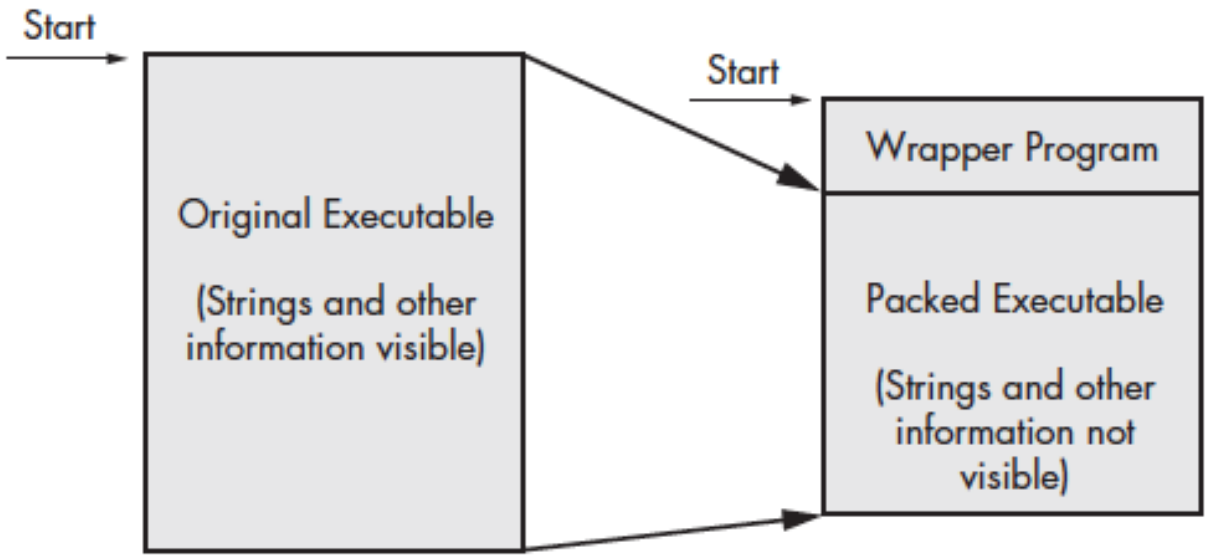
Packing or Obfuscation

- Malware writers often use *packing* or *obfuscation* to make their files more difficult to detect or analyze.
- *Obfuscated* programs are ones whose execution the malware author has attempted to hide.
- *Packed* programs are a subset of obfuscated programs in which the malicious program is compressed and cannot be analyzed.
- *Legitimate* programs almost always include many strings.
  - Malware that is packed or obfuscated contains very few strings. If upon searching a program with Strings,
  - you find that it has only a few strings, it is probably either obfuscated or packed,

Packing Files I

Packing Files

- When the packed program is run, a small wrapper program also runs to decompress the packed file and then run the unpacked file.

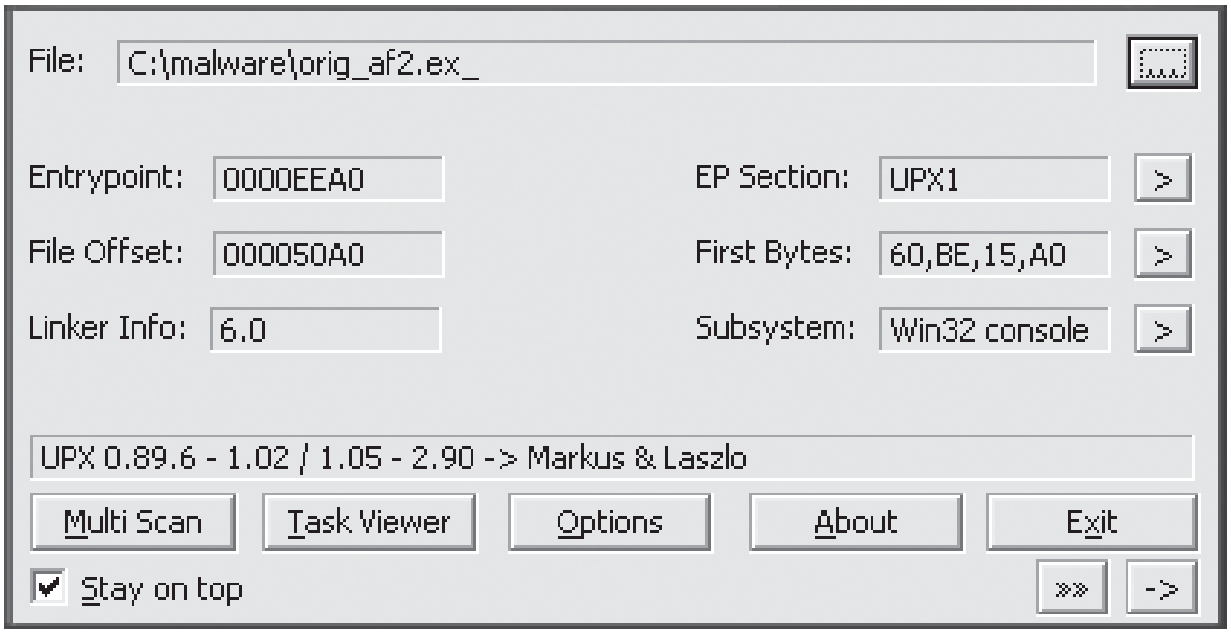


Şekil: The file on the left is the original executable, with all strings, imports, and other information visible. On the right is a packed executable. All of the packed file's strings, imports, and other information are compressed and invisible to most static analysis tools.

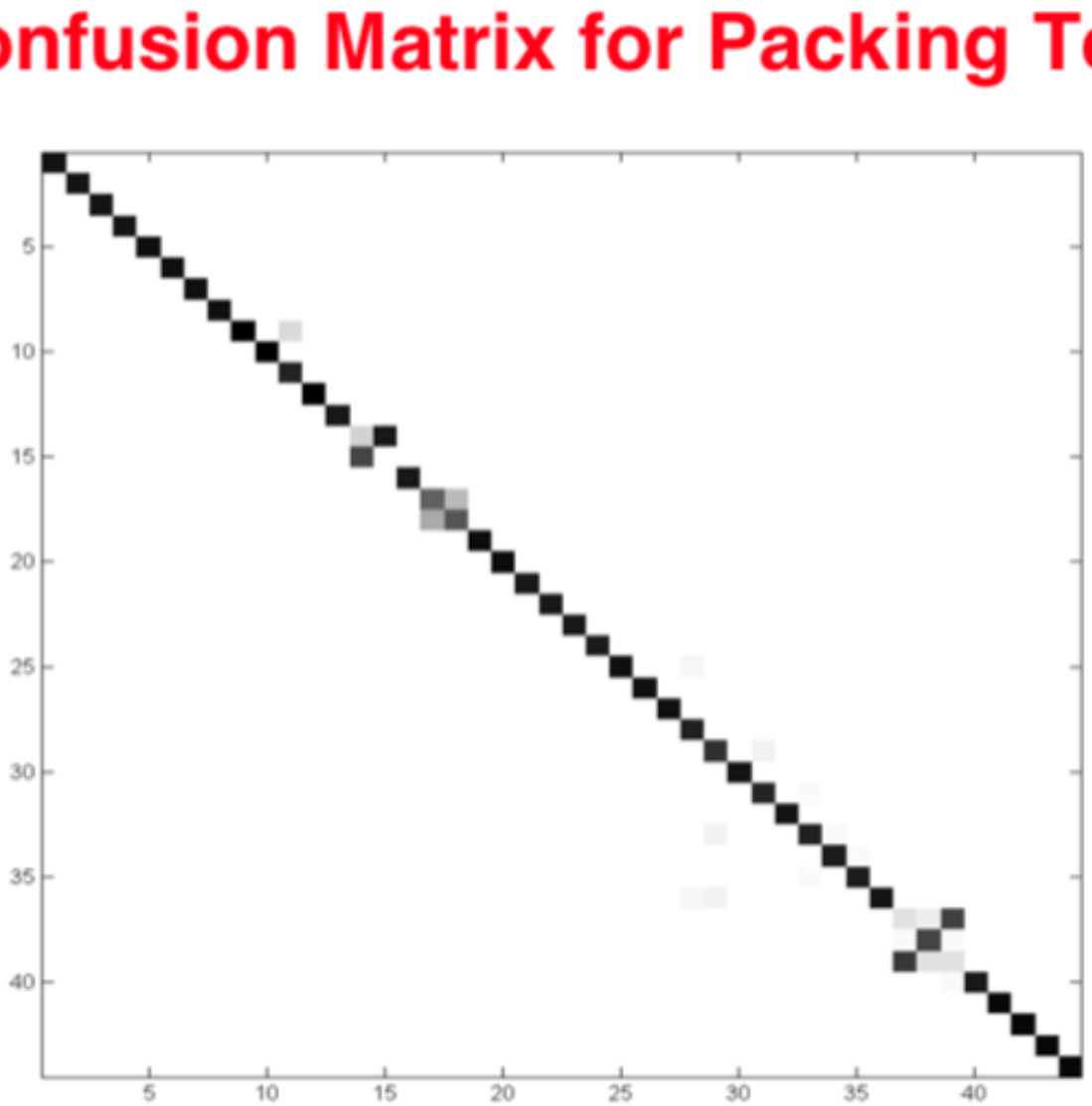
Packing Files II

PEiD

- One way to detect packed files is with the PEiD program.
- You can use PEiD to detect the type of packer or compiler employed to build an application, which makes analyzing the packed file much easier.
- As you can see, PEiD has identified the file as being packed with UPX version 0.89.6-1.02 or 1.05-2.90
- When a program is packed, you must unpack it in order to be able to perform any analysis.



Packing Confusion Matrix



İçindekiler

Biyometrik Kimlik Doğrulama I

Touchalytics

- Touchscreen kullanılarak kimlik doğrulaması
- Wikipedia sayfalarının (Wind, Tulip Mania, Yosemite National Park) okunması sırasında kullanıcıların ekranda yaptığı işlemler
- Raw features: finger up, finger down, finger move, multi-touch

Biyometrik Kimlik Doğrulama II