# Anwendbarkeit der Entropie-Analyse für die Erkennung von Shellcode

Michael Gröning

DFN-CERT Services GmbH
HAW Hamburg

DFN CERT

# Einleitung
## Kurze Vorstellung

- Studium der Technischen Informatik(B.Sc.) an der HAW-Hamburg
- Arbeit beim DFN-CERT seit 2006
- Incident Response Team - Aufgabe: Schwachstellenanalyse und Bewertung
- Bachelorarbeit: Üntersuchung von unbekannten Angriffen auf Low-Interaction Honeypots"

# Einleitung
## Inhalte dieses Vortrages

- Einführung in die Problemstellung
- Begriffserklärung: Entropie
- Anwendung der Entropie zur Schadcode-Erkennung
- Ergebnisse von ersten Tests

# Aufgabe

Weiterentwicklung eines Low-Interaction Honeypots für die gezielte Untersuchung von bisher nicht bekannten Phänomen.

# Problemstellung

- Entwicklung heuristischer Verfahren für die Klassifizierung erhobener Daten
- Zero-Knowledge! Es werden vorher keine Annahmen über die Daten getroffen (z.B. basierend auf TCP/UDP-Port, verwendetem Protokoll etc.).
- Die Verfahren sollen effizient und vor allem sehr schnell an veränderte Situationen anpassbar sein.
- Ergänzung zu bereits vorhandenen Low-Interaction Honeypots wie z.B. Amun[1] oder Nepenthes[2], kein Ersatz!

[1] http://amunhoney.sourceforge.net
[2] http://nepenthes.carnivore.it

DFN CERT

# Problemstellung

- Entwicklung heuristischer Verfahren für die Klassifizierung erhobener Daten
- Zero-Knowledge! Es werden vorher keine Annahmen über die Daten getroffen (z.B. basierend auf TCP/UDP-Port, verwendetem Protokoll etc.).
- Die Verfahren sollen effizient und vor allem sehr schnell an veränderte Situationen anpassbar sein.
- Ergänzung zu bereits vorhandenen Low-Interaction Honeypots wie z.B. Amun[1] oder Nepenthes[2], kein Ersatz!

[1] http://amunhoney.sourceforge.net
[2] http://nepenthes.carnivore.it

DFN CERT

# Problemstellung

- Entwicklung heuristischer Verfahren für die Klassifizierung erhobener Daten
- Zero-Knowledge! Es werden vorher keine Annahmen über die Daten getroffen (z.B. basierend auf TCP/UDP-Port, verwendetem Protokoll etc.).
- Die Verfahren sollen effizient und vor allem sehr schnell an veränderte Situationen anpassbar sein.
- Ergänzung zu bereits vorhandenen Low-Interaction Honeypots wie z.B. Amun[1] oder Nepenthes[2], kein Ersatz!

[1] http://amunhoney.sourceforge.net
[2] http://nepenthes.carnivore.it

DFN CERT

# Problemstellung

- Entwicklung heuristischer Verfahren für die Klassifizierung erhobener Daten
- Zero-Knowledge! Es werden vorher keine Annahmen über die Daten getroffen (z.B. basierend auf TCP/UDP-Port, verwendetem Protokoll etc.).
- Die Verfahren sollen effizient und vor allem sehr schnell an veränderte Situationen anpassbar sein.
- Ergänzung zu bereits vorhandenen Low-Interaction Honeypots wie z.B. Amun[1] oder Nepenthes[2], kein Ersatz!

[1] http://amunhoney.sourceforge.net
[2] http://nepenthes.carnivore.it

DFN CERT

# Einführung
## Abgrenzung

- Das Hauptaugenmerk liegt vor allem auf bisher unbekannten und kurzlebigen Phänomenen.
- keine ausgeklügelte Kommunikation mit dem potentiellen Angreifer.
- keine Untersuchung von Standard-Ports, die von Botnetzen verwendet werden (Port 445, 139 etc).

DFN CERT

# ähnliche Arbeiten

Robert Lyda und James Hamrock:
*Using Entropy Analysis to Find Encrypted and Packed Malware.*
*IEEE Security and Privacy 5 (2007)*

- Suche von Malware in Windows Binaries.
- Hauptaugenmerk auf verschlüsselten und komprimierten Binaries

DFN CERT

# ähnliche Arbeiten

Paul Bächer und Markus Koetter:
LibEmu[3] - x86 shellcode detection and emulation

- Emulation einer x86 CPU
- Emulation von Teilen der Windows API
- Ermöglicht die Shellcode-Erkennung und Ausführung.

[3] http://libemu.carnivore.it

# Entropie

## Was ist Entropie?

Entropie ist definiert als das Maß für den mittleren Informationsgehalt pro Zeichen eines Textes.

# Entropie
## Was ist Entropie?

Entropie ist definiert als das Maß für den mittleren Informationsgehalt pro Zeichen eines Textes.

# Entropie

## Definition

Sei $\Sigma = \{a_1, a_2, \ldots, a_n\}$ ein endliches Alphabet und $x \in \Sigma^*$, in dem das Zeichen $a_i$ mit der Wahrscheinlichkeit $p_i$ auftritt. Die Entropie $H : \Sigma^* \rightarrow \mathbb{R}$ von $x$ ist:

$$H(x) = -\sum_{i=1}^{n} p(i) \log_2 p(i)$$

DFN CERT

# Entropie

$$H(x) = -\sum_{i=1}^{n} p(i) \log_2 p(i)$$

Wird eine Nachricht betrachtet, so tritt jedes Zeichen $i$ aus dem beschreibenden Zeichenalphabet $\Sigma$mit einer bestimmten Wahrscheinlichkeit $p(i)$ bezogen auf die Gesamtanzahl der Zeichen auf.
Die aus der Summe dieser Wahrscheinlichkeiten berechnete Entropie, kann für verschiedene Typen von Nachrichten charakteristisch sein.

DFN CERT

# Entropie
## Beispiele

- Münzwurf, Würfel: Gleichverteilung der Wahrscheinlichkeiten.
- Datenkomprimierung: Entropie als Maß für die Komprimierbarkeit von Daten.

# Entropie
## Beispiele

- Münzwurf, Würfel: Gleichverteilung der Wahrscheinlichkeiten.
- Datenkomprimierung: Entropie als Maß für die Komprimierbarkeit von Daten.

# Entropie
## Beispiele

- Münzwurf, Würfel: Gleichverteilung der Wahrscheinlichkeiten.
- Datenkomprimierung: Entropie als Maß für die Komprimierbarkeit von Daten.

# Entropie-Analyse

Eigenschaften von Shellcode:

- Shellcode muss i. Allg. in den übertragenen Nutzdaten enthalten werden.(Unterschiedliche Struktur/Entropie?)
- Shellcode besteht aus Assemblercode.
- Die Nutzdaten weisen häufig Auffälligkeiten auf (z.B. überlange Datenfelder, NOP-Sleds,...)

DFN CERT

# Entropie-Analyse

Eigenschaften von Shellcode:

- Shellcode muss i. Allg. in den übertragenen Nutzdaten enthalten werden.(Unterschiedliche Struktur/Entropie?)
- Shellcode besteht aus Assemblercode.
- Die Nutzdaten weisen häufig Auffälligkeiten auf (z.B. überlange Datenfelder, NOP-Sleds,...)

DFN CERT

# Entropie-Analyse

Eigenschaften von Shellcode:

- Shellcode muss i. Allg. in den übertragenen Nutzdaten enthalten werden.(Unterschiedliche Struktur/Entropie?)
- Shellcode besteht aus Assemblercode.
- Die Nutzdaten weisen häufig Auffälligkeiten auf (z.B. überlange Datenfelder, NOP-Sleds,...)

DFN CERT

# Entropie-Analyse

Eigenschaften von Shellcode:

- Shellcode muss i. Allg. in den übertragenen Nutzdaten enthalten werden.(Unterschiedliche Struktur/Entropie?)
- Shellcode besteht aus Assemblercode.
- Die Nutzdaten weisen häufig Auffälligkeiten auf (z.B. überlange Datenfelder, NOP-Sleds,...)

DFN CERT

# Entropie-Analyse
## Vorgehensweise

### Beispiel:

- Ein anderes System baut eine Verbindung zum Honeypot auf.
- Von den empfangenen Daten wird die globale Entropie sowie mit Hilfe eines Sliding-Window Algorithmus eine lokale Entropie bestimmt.
- Pakete die deutliche Unterschiede zwischen globaler und maximaler Entropie aufweisen, werden näher betrachtet.

DFN CERT

# Entropie-Analyse
## Vorgehensweise

### Beispiel:

- Ein anderes System baut eine Verbindung zum Honeypot auf.
- Von den empfangenen Daten wird die globale Entropie sowie mit Hilfe eines Sliding-Window Algorithmus eine lokale Entropie bestimmt.
- Pakete die deutliche Unterschiede zwischen globaler und maximaler Entropie aufweisen, werden näher betrachtet.

DFN CERT

# Entropie-Analyse
## Vorgehensweise

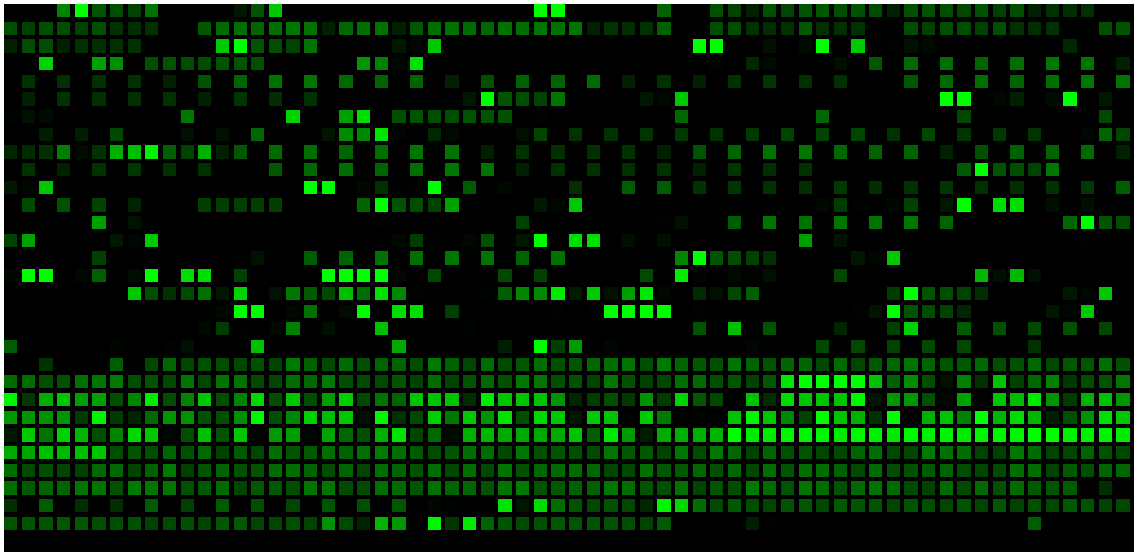
### Beispiel:

- Ein anderes System baut eine Verbindung zum Honeypot auf.
- Von den empfangenen Daten wird die globale Entropie sowie mit Hilfe eines Sliding-Window Algorithmus eine lokale Entropie bestimmt.
- Pakete die deutliche Unterschiede zwischen globaler und maximaler Entropie aufweisen, werden näher betrachtet.

DFN CERT

# Entropie-Analyse
## Visualisierung des Paketes

Graphische Darstellung eines Angriffes auf den Honeypot:

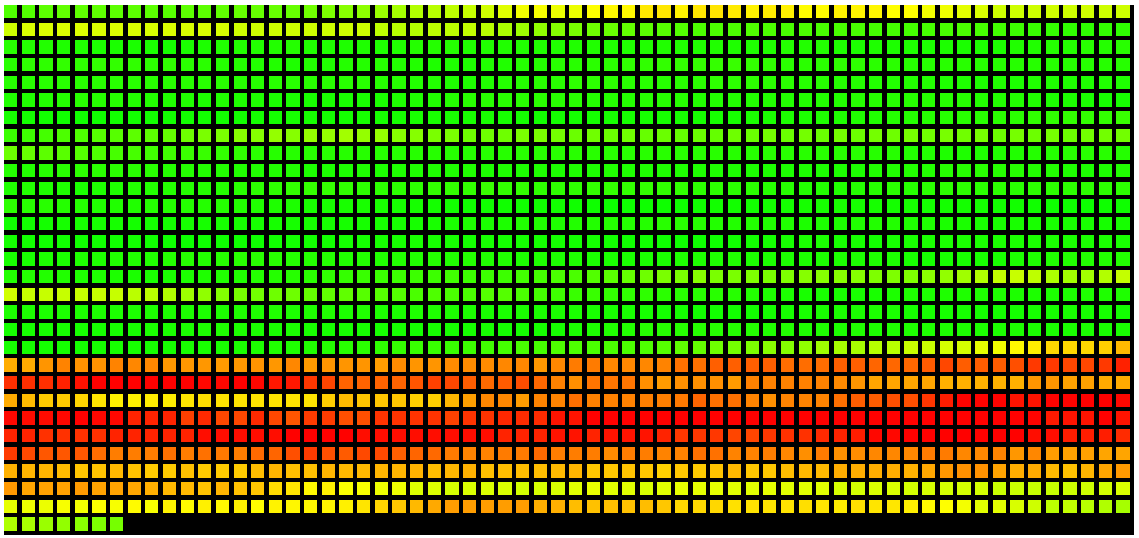


Darstellung des Bytestreams eines Paketes (0x00 = Schwarz, 0xFF = Hellgrün)

# Entropie-Analyse

## Visualisierung der lokalen Entropiewerte

Graphische Darstellung der Verschiedenen Entropiewerte des Sliding Windows



Starker Anstieg der Entropie im hinteren Drittel der untersuchten Daten.

# Entropie-Analyse

## Vergleich zwischen Entropie-Analyse und CPU-Emulation

Untersuchung mit Hilfe des Tools sctest[4]:

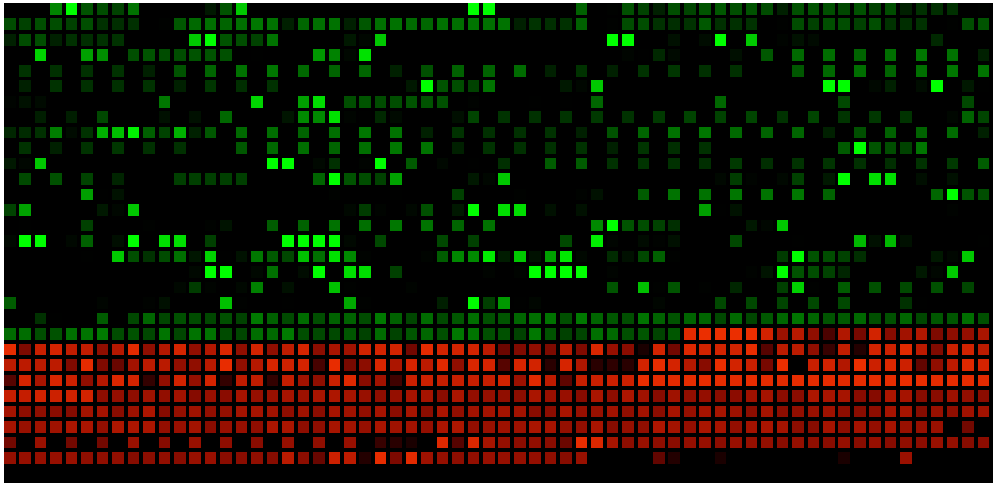
```
$ ./bin/sctest -gS -s 1000000  < data/445.dump
verbose = 0
success offset = 0x0000056c
Hook me Captain Cook!
userhooks.c:132 user_hook_ExitThread
ExitThread(0)
stepcount 85071
HMODULE LoadLibraryA (
     LPCTSTR lpFileName = 0x0041761a =>
           = "urlmon";
...
```

[4] http://libemu.carnivore.it

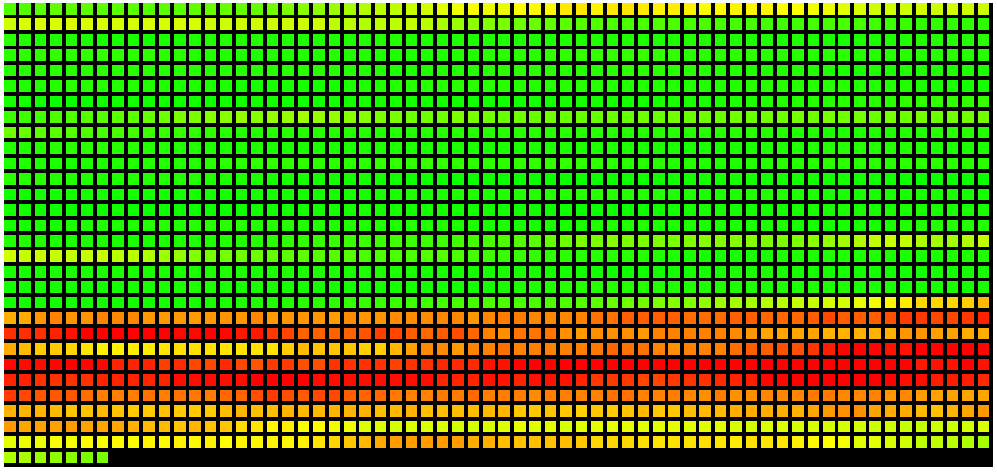
# Entropie-Analyse

## Vergleich zwischen Entropie-Analyse und CPU-Emulation anhand der Visualisierung

Graphische Darstellung mit Rot eingefärbten Shellcodeblock



Graphische Darstellung der lokalen Entropiewerte des Sliding Window



DFN CERT

# Entropie-Analyse

## Vergleich zwischen Entropie-Analyse und CPU-Emulation

Anzahl der Angriff klassifizierten Verbindungsversuche auf Verschiedene TCP-Ports.

| TCP-Port | Entropie-Analyse | libEmu |
|---|---|---|
| 1023 | 10 | 10 |
| 1085 | 0 | 2 |
| 1087 | 2 | 2 |
| 1096 | 1 | 1 |
| 2967 | 5095 | 5095 |
| 3050 | 18 | 2135 |
| 5554 | 29 | 29 |
| 7986 | 0 | 2 |
| 8800 | 0 | 43 |
| 41523 | 97 | 97 |

Zeitraum: 06-17-2009 - 09-07-2009
ca. 350.000 Untersuchte Verbindungen
Parameter:

Entropie: avg <=5, max >=5 ; libEmu: GETPC only

DFN CERT

# Entropie-Analyse
## Fazit

- Anwendung des Verfahrens sehr gut geeignet für Low-Interaction Honeypots
- Einfache Implementierung gegenüber CPU-Emulation
- Sehr schnelles Verfahren

DFN CERT

# Entropie-Analyse
## Fazit

- Anwendung des Verfahrens sehr gut geeignet für Low-Interaction Honeypots
- Einfache Implementierung gegenüber CPU-Emulation
- Sehr schnelles Verfahren

# Entropie-Analyse
## Fazit

- Anwendung des Verfahrens sehr gut geeignet für Low-Interaction Honeypots
- Einfache Implementierung gegenüber CPU-Emulation
- Sehr schnelles Verfahren

# Fragen?