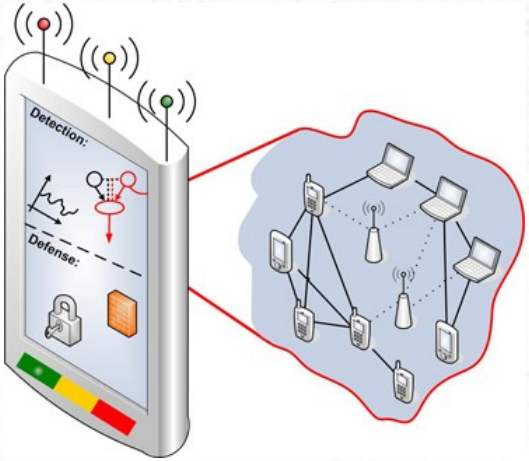
AW2 - Leichtgewichtige Schadcodeerkennung

als Teil des SKIMS Projekts

Benjamin Jochheim
Betreuer: Thomas Schmidt



iNET
Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Gliederung

- Aktueller Stand
- Ziel der Arbeit
- Entropie, Frequenzanalyse
- Verwandte Arbeiten
  - Net-Entropy
  - OSCAR-Methode
  - Sliding Window Methode
  - Malware Analyse via Statistischer Methoden
- Abgrenzung
- Ausblick

Rückblick / Aktueller Stand

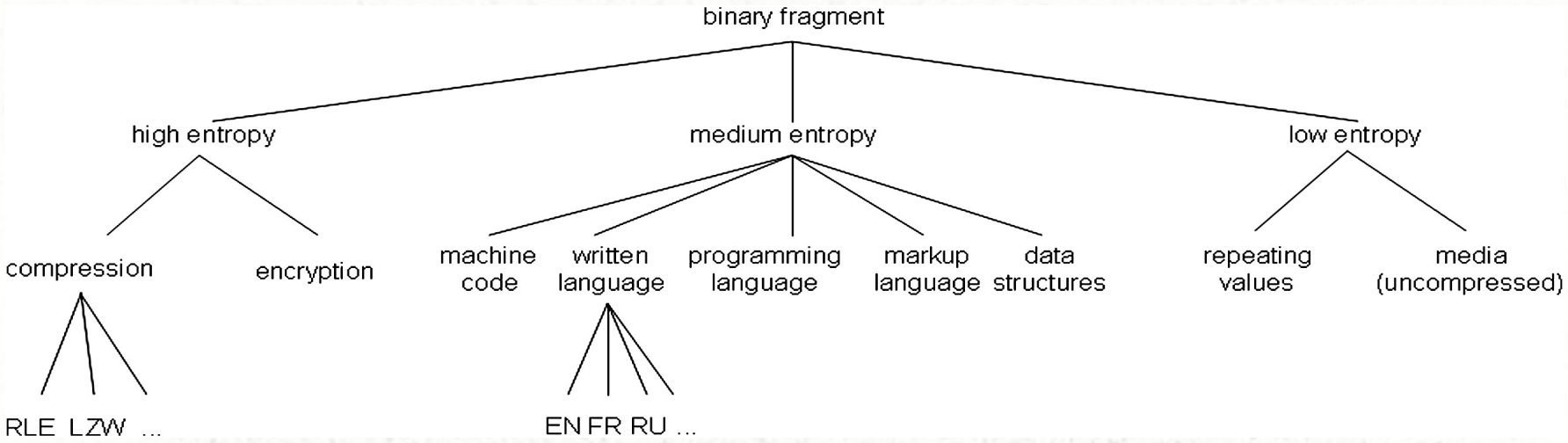
- Themenwechsel
  - ~~Sichere Gruppenkommunikation~~
  - Leichtgewichtige Schadcodeerkennung
    - Teil des SKIMS Projekts
    - Terena 2011 student poster competition
    - Poster angenommen zur WiSec´11
    - MATLAB Prototyp

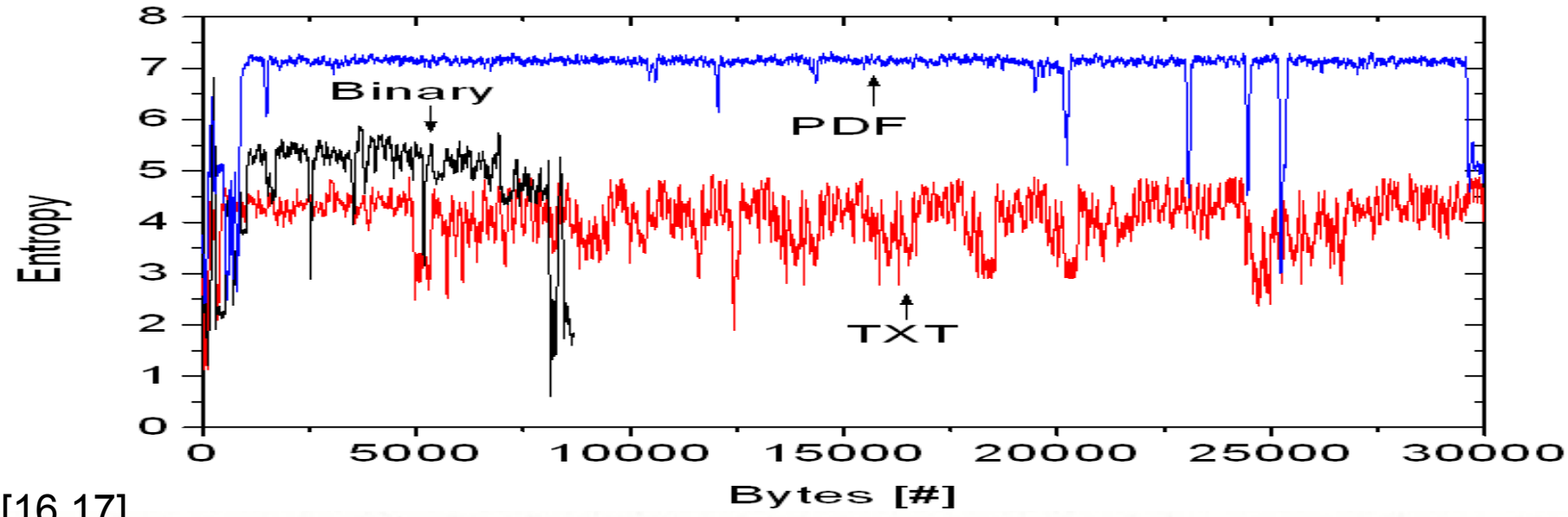
« networking the networkers »

Ziel der Arbeit

- Problem: Erkennung von Attacken auf Mobiltelefonen
- Binary Instruction Code VS. Regulärer Code
- Zero-Knowledge Ansatz
- leichtgewichtig
- Techniken
  - Entropie
  - Frequenzanalyse

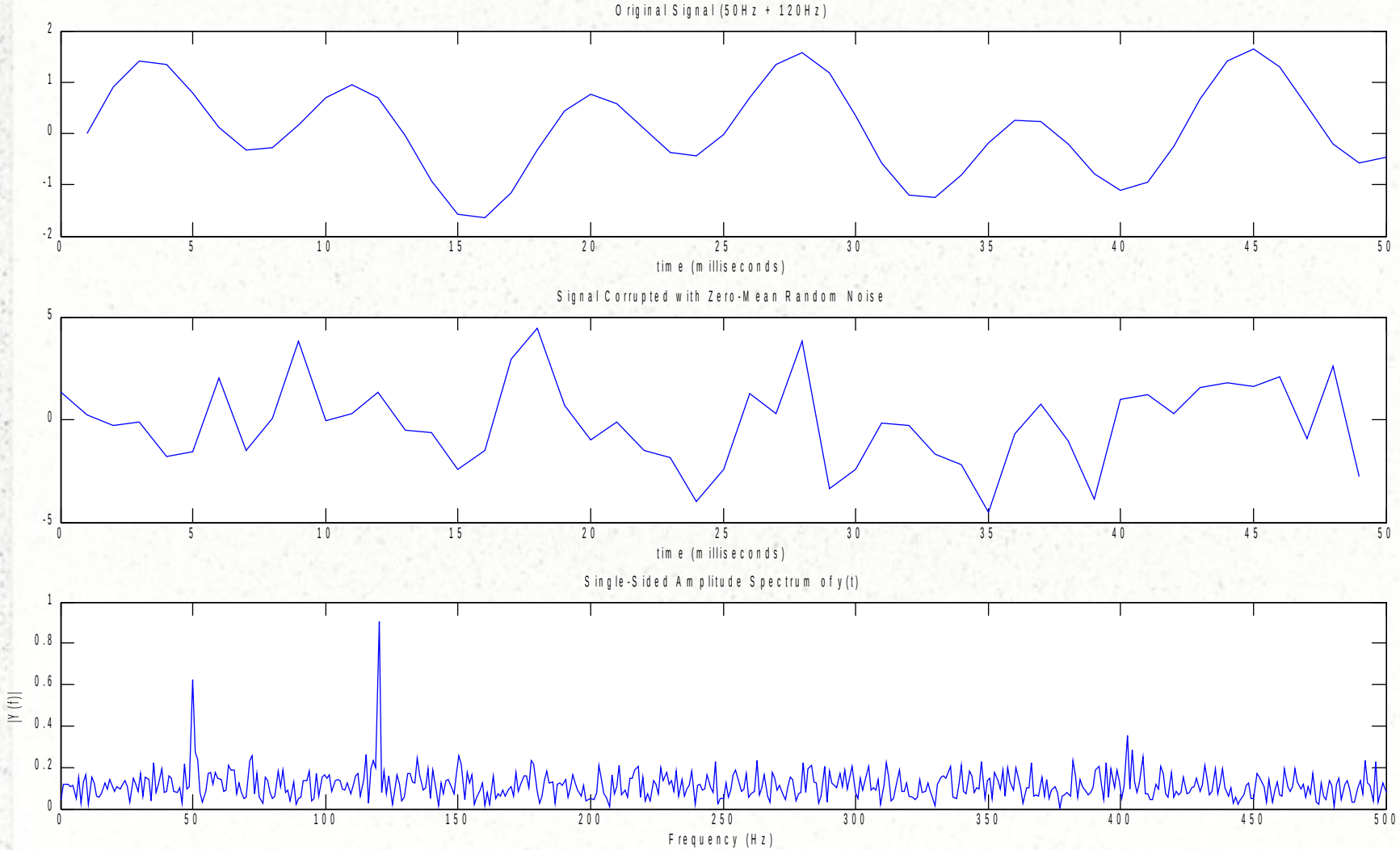
Entropie





[16,17]

Frequenzanalyse



Verwandte Arbeiten eine Übersicht

- Schwierige Vergleichbarkeit
- Statische Schadcodeerkennung
  - Traditionelle Malware-Erkennung
    - Syntaktische Ansätze
      - Prüfsummen
    - Semantische Ansätze [10]
  - Data-Mining Ansätze [13,14]

Verwandte Arbeiten eine Übersicht (2)

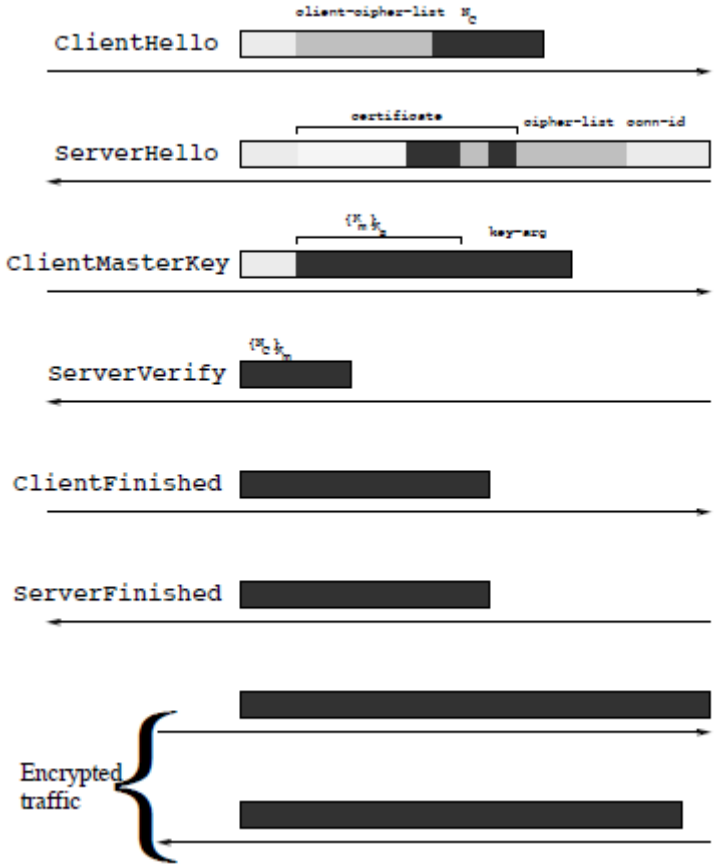
- Leichtgewichtige Malware-Erkennung [5,9,12]
- Computer Forensik [11]
  - Byte Frequenzverteilung (BFD), Histogramme
  - Metrik Ansätze (z.B. Entropie)
  - Kombination der beiden obigen (BFD für Grob-Erkennung)

NET-Entropy

Jean Goubault-Larrecq and Julien Olivain

2006

- Erkennung von Attacken in verschlüsselten Datenströmen
  - z.B. SSL, SSH
- Beginn der Kommunikation
- Entropiebasiert
  - Mit Entropieschätzer [4]
- Entropieverteilung
- Quelle: [5]



SSL Sessionaufbau, [5]

OSCAR-Methode

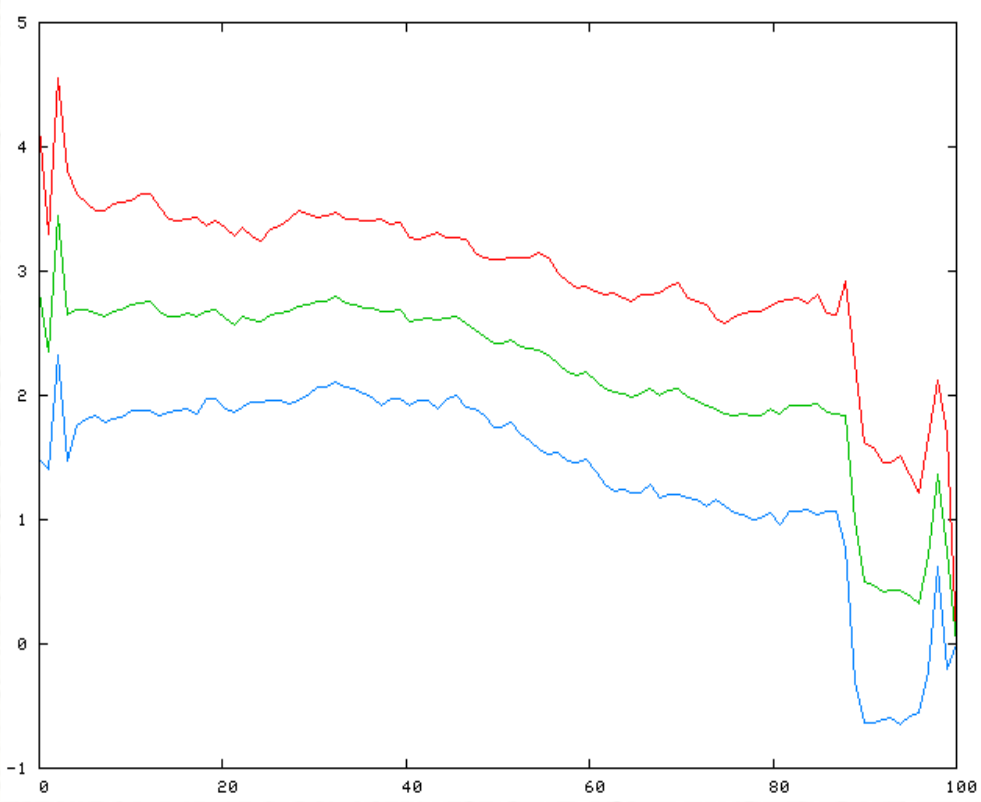
Karresand et.al.

2006

- Byte frequency distribution (BFD) basiert
- RoC „Rate of Change“ Metrik
- Erkennung mittels Distanzmetrik zu geclusterten Vergleichsdaten auf 4kB Abschnitten
- Funktioniert gut mit JPEG-Dateien (99.2% Erkennungsrate) wg. gut erkennbaren Strukturen [6]
- Quelle: [2,3]

Sliding Window Measurement for File Type Identification
Gregory A. Hall et.al.
2006

- Entropie und komprimierbarkeit
- sliding window
- Metrik:
  - point-by-point delta
  - Pearson's Rank Order Correlation
- Quelle:[8]



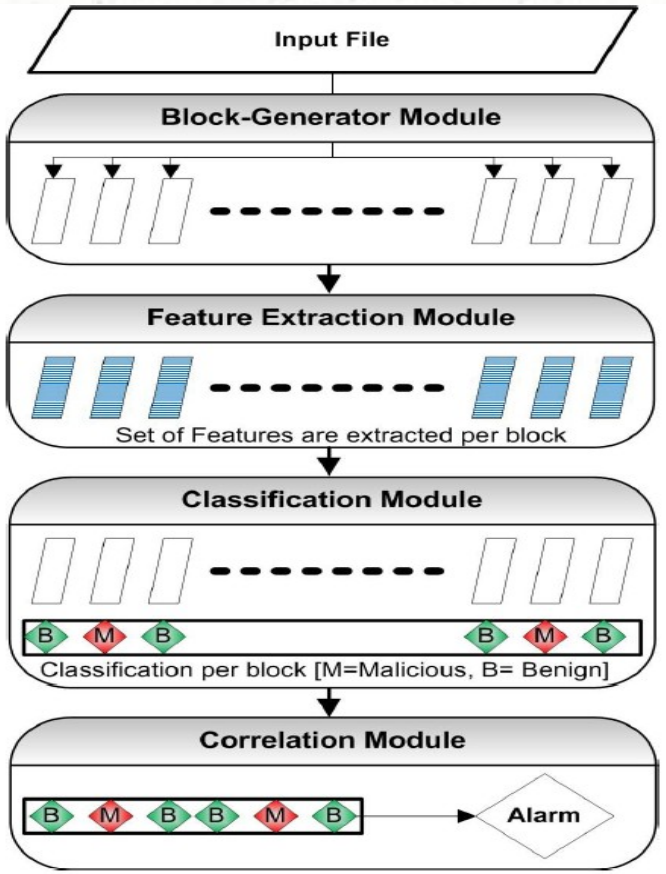
Entropiefunktion einer PE-EXE-Datei

Malware Detection using Statistical Analysis of Byte-Level File Content

S. Momina Tabesh et.al.

2009

- Block weises Vorgehen
- 13 Statistische Messgrößen
  - Pro 1-,2-,3-,4- Gram der Blöcke
- Klassifikator:
  - Entscheidungsbäume [15]
- Quelle:[12]



Abgrenzung zu anderen Arbeiten

- Zero Knowledge Verfahren
  - Entropie & Frequenzanalyse
- Leichtgewichtig
- Identifikation kleiner Abschnitte in Dateien
- Keine Dateiidentifikation sondern nur binary VS. nonbinary unterscheiden
- Verfahren evtl. später mit Kenntnis des Dateityps
- Probleme:
  - Native Code VS. Dalvik Code
  - komprimierter Schadcode

Ausblick

- Experimente mit Fenstergrößen
- Entropieschätzfunktion nutzen
- Wie leichtgewichtig ist das Verfahren?
  - Prototyp auf Android Platform umsetzen
- Testdatensätze weiter ausbauen
- WiSec 2011 Poster

Vielen Dank für eure Aufmerksamkeit!
Fragen !?!

Referenzen

- **[1]** Li WJ, Wang K, Stolfo SJ, Herzog B. “Fileprints: identifying file types by n-gram analysis”. 6th IEEE Information Assurance Workshop, West Point, NY, June, 2005
- **[2]** Karresand M, Shahmehri N. “Oscar—file type identification of binary data in disk clusters and RAM pages,” in Proceedings of IFIP International Information Security Conference: Security and Privacy in Dynamic Environments (SEC2006), LNCS, p413-424. 2006.
- **[3]** Karresand M, Shahmehri N. “File Type Identification of Data Fragments by Their Binary Structure”, Proceedings of the 7th Annual IEEE Information Assurance Workshop, "The West Point Workshop", pp 140-147, United States Military Academy, West Point, 21-23, New York, June 2006
- **[4]** IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 50, NO. 9, SEPTEMBER 2004, Liam Paninski, Estimating Entropy on Bins Given Fewer Than Samples, http://www.stat.columbia.edu/~liam/research/abstracts/nm_proof-abs.html

Referenzen(2)

- **[5]** Jean Goubault-Larrecq and Julien Olivain, Detecting Subverted Cryptographic Protocols by Entropy Checking, 2006, http://www.lsv.ens-cachan.fr/Publis/RAPPORTS_LSV/PDF/rr-lsv-2006-13.pdf
- **[6]** Roussev, Vassil, and Garfinkel, Simson, "File Classification Fragment-The Case for Specialized Approaches," Systematic Approaches to Digital Forensics Engineering (IEEE/SADFE 2009), Oakland, California
- **[7]** Calhoun WC, Coles D. “Predicting the types of file fragments.” Proceedings of the 2008 DFRWS Conference, Baltimore, MD. Aug 2008. pp.146-157.
- **[8]** Gregory A. Hall, Sliding Window Measurement for File Type Identification, Computer Forensics and Intrusion Analysis Group, ManTech Security and Mission Assurance, 2006.

Referenzen(3)

- **[9]** M. Weber et al., “A Toolkit for Detecting and Analyzing Malicious Software,” Proc. 18th Ann. Computer Security Applications Conf., IEEE CS Press, 2002, pp. 423–431.
- **[10]** Mihai Christodorescu et. al., „Semantics-Aware Malware Detection“, IEEE Symposium on Security and Privacy, Oakland, California, May 2005
- **[11]** Zeitliche Auflistung relevanter Forensik-Arbeiten zur Dateifragmenterkennung (Abrufdatum: 1. Juni 2011) http://www.forensicswiki.org/wiki/File_Format_Identification
- **[12]** S. Momina Tabish, M. Zubair Shafiq, and Muddassar Farooq. 2009. Malware detection using statistical analysis of byte-level file content. In Proceedings of the ACM SIGKDD Workshop on CyberSecurity and Intelligence Informatics (CSI-KDD '09)
- **[13]** Jeremy Z. Kolter and Marcus A. Maloof. 2004. Learning to detect malicious executables in the wild. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '04).

Referenzen(4)

- **[14]** M.G. Schultz, E. Eskin, E. Zadok, S.J. Stolfo, "Data mining methods for detection of new malicious executables", IEEE Symposium on Security and Privacy, pp. 38-49, USA, IEEe Press, 2001.
- **[15]** Y. Freund, R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting", Journal of Computer and System Sciences, No. 55, pp. 23-37, 1997
- **[16]** T. Schmidt, M.Waehlisch, M.Groening, „Context-adaptive Entropy Analysis as a Lightweight Detector of Zero-day Shellcode Intrusion for Mobiles", WiSec Conference 2011
- **[17]** „Automated mapping of large binary objects using primitive fragment type classification", Science Direct, digital investigation 7; S 3.S12, 2010