Frühe Warnung durch Beobachten und Verfolgen von bösartiger Software im Deutschen Internet: *Das Internet-Malware-Analyse System* (InMAS)

Markus Engelberth[1], Felix Freiling[1], Jan Göbel[1], Christian Gorecki[1],

Thorsten Holz[1], Philipp Trinius[1], Carsten Willems[1]

Kurzfassung:
Das frühzeitige Erkennen neuer Gefahren und Anomalie im Internet ist Ziel eines *Frühwarnsystems*. Ein wichtiger Aspekt eines solchen Systems ist die Beobachtung und Verfolgung von bösartiger Software, da diese Angriffstechnik zu einem großen Teil für aktuelle Bedrohungen im Internet verantwortlich ist. In diesem Beitrag wird das *Internet-Malware-Analyse-System* (InMAS) vorgestellt, mit dessen Hilfe bösartige Software automatisiert gesammelt, analysiert und ausgewertet werden kann. Die so gewonnenen Informationen liefern ein Lagebild zum aktuellen Gefährdungslevel im Internet und liefern somit einen wichtigen Beitrag für ein nationales Frühwarnsystem.

Stichworte: Schadsoftware, Frühwarnungsystem, Internet

1. Einführung

1.1 Motivation

Das Internet zählt heute zu den kritischen Infrastrukturen der Bundesrepublik Deutschland. Ein (auch nur teilweiser) Ausfall dieses Netzes hätte dramatische Auswirkungen auf Wirtschaft und Gesellschaft. Trotz einer robusten Auslegung der Protokolle und der Netzinfrastruktur hat es immer wieder Vorfälle gegeben, in denen die Verfügbarkeit des Internets merklich reduziert war − beispielsweise Code Red im Jahr 2001 (Moore, Shannon, & Claffy, 2002). Derartige Vorfälle zu antizipieren und möglichst frühzeitig zu erkennen ist das Ziel eines *Frühwarnsystems*. Die Warnungen dieses Systems bilden die Grundlage für eine gezielte und wirkungsvolle Reaktion auf drohende Angriffe.

Ein Frühwarnsystem sollte zwei Funktionen erfüllen: Zum einen soll es dem Betreiber einen Einblick in den momentanen Zustand des Netzes geben und somit eine globale Sicht auf das Gesamtsystem liefern. Zum anderen soll es Anomalien im Netz automatisch erkennen und möglichst genau klassifizieren können. Von besonderem Interesse sind hierbei gezielte und massive Angriffe auf die Netzinfrastruktur und damit verbundenen Systemausfälle. Angriffe dieser Art werden heute zumeist durch bösartige Software (*malicious software*, kurz *Malware*) verursacht, die sich autonom verbreitet und dabei in großem Maßstab Rechner im Internet infiziert. Beispiele für derartige Malware sind Würmer, Trojanische Pferde und Bots. Insbesondere der Zusammenschluss von Bots zu großen Botnetzen stellt eine ernstzunehmende Gefahr

[1] Lehrstuhl für Praktische Informatik 1, Universität Mannheim

für das Internet dar. Ein Frühwarnsystem sollte darum stets einen Überblick über den aktuellen Gefährdungsgrad durch Malware liefern.

1.2 Honeypots und Honeynets

Ein *Honeypot* ist ein Netzwerkelement (Rechner, Router, Switch), der ans Internet angeschlossen wird, um angegriffen und kompromittiert zu werden. Diese Form von elektronischen Ködern eignet sich besonders gut, um sich autonom verbreitende Malware zu sammeln. Der Honeypot simuliert hierzu die Schwachstellen verwundbarer Netzwerkdienste und kopiert bei einer simulierten Kompromittierung den Schadcode in eine Datenbank. Der Schadcode kann anschließend automatisiert analysiert werden. Mehrere Honeypots können zu einem Netz zusammengeschlossen werden, um so eine größere Menge und Vielfalt von Malware anzulocken. Ein solches Netz von Honeypots nennt man *Honeynet*.

Im Rahmen eines Frühwarnsystems können Honeypots und Honeynets als Teil der Sensorik eingesetzt werden, um frühzeitig Informationen zu neuen Sicherheitsvorfällen zu sammeln.

1.3 Internet–Malware –Analyse–System (InMAS)

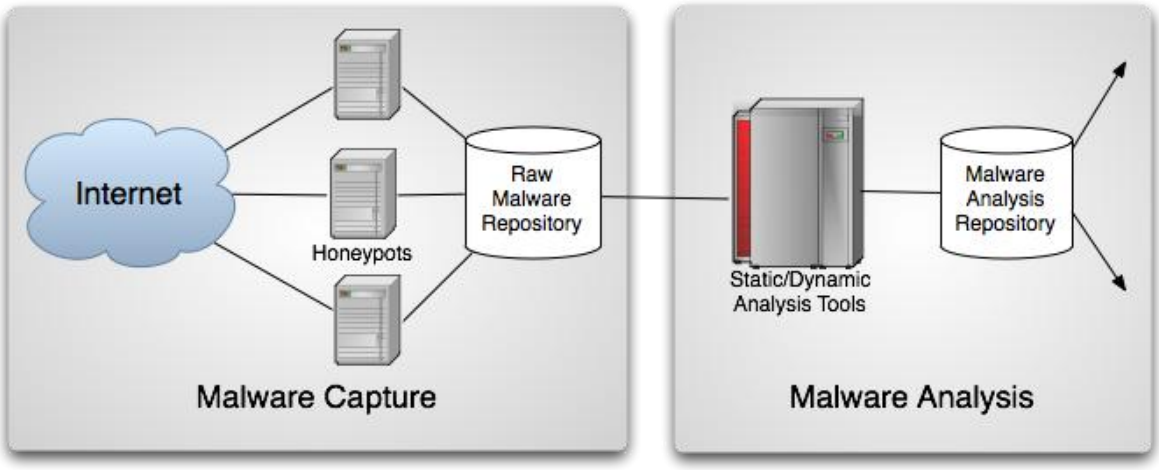

Abbildung 1: Schematische Übersicht zu InMAS

Das Internet-Malware-Analyse-System (InMAS) ist ein verteiltes Sensor-System zur Sammlung und Analyse von Malware. InMAS basiert im Kern auf der Honeypot-Software Nepenthes (Bächer, Freiling, Holz, Dornseif, & Kötter, 2006) und dem Analysetool CWSandbox (Willems, Freiling, & Holz, 2007), die beide an der Universität Mannheim weiterentwickelt werden. InMAS erweitert das nationale Frühwarnsystem um die Fähigkeit, Malware in allen seinen Erscheinungsformen zu sammeln, detailliert zu analysieren, sowie die Informationen über die gesammelte Malware mit anderen Datenquellen oder Analysesystemen zu koppeln. Abbildung 1 gibt einen schematischen Überblick über den Aufbau von InMAS.

InMAS lässt sich in die zwei Teilsysteme *Malware Capture* und *Malware Analysis* untergliedern. Diese Trennung zwischen Senorik und Auswertung wurde in erster Linie aus Gründen der Skalierbarkeit vorgenommen. Des Weiteren erlaubt der modulare Aufbau − dieser setzt sich auch innerhalb der Teilsysteme fort − eine einfache Erweiterung und Wartung des Systems. Beide Teilsysteme verfügen über eine eigene Datenbank:

- Das *Raw Malware Repository* nimmt die gefundene Malware auf. Die Malware wird hier „roh“, das heißt ohne jegliche Analyseergebnisse abgelegt. Dieses Repository ist die Grundlage für weitere Auswertungsschritte und bildet die Schnittstelle zwischen der Sensorik und dem Auswerungssystem.
- In dem *Malware Analysis Repository* wird die Malware in einer mit verschiedenen Informationen angereicherten Art und Weise abgespeichert. Hierzu gehören Analyseergebnisse der CWSandbox, von kommerziellen Virenscannern und Metainformationen. Von diesem Repository aus gibt es mehrere Schnittstellen zu weiteren Auswertesystemen wie beispielsweise dem Internet-Analyse-System (IAS) oder zum Lagezentrum, in dem Analysten dann auf die gesammelten Informationen zugreifen können.

Die folgenden beiden Abschnitte beschreiben ausgewählte Komponenten der Sensorik und des Auswerungssystem von InMAS genauer. Im Anschluß daran werden in Abschnitt 4 beispielhaft einige der mittels InMAS gesammelten Informationen vorgestellt. Der Beitrag endet in Abschnitt 5 mit einer Zusammenfassung und einer Übersicht zu zukünftigen Arbeiten in diesem Bereich.

2. InMAS Sensorik (*Malware Capture*)

InMAS verfügt in der aktuellen Konfiguration über drei verschiedene Sensor/ Honeypot-Typen. Zum Sammeln „klassischer“ Malware, darunter fallen beispielsweise Würmer und Bots die sich über bekannte Schwachstellen in Netzwerkdiensten verbreiten, kommt die Honeypot-Lösung Nepenthes zum Einsatz. Die an diesem System vorgenommenen Erweiterung werden in Abschnitt 2.1 erläutert. Einer der größten Schwächen von Nepenthes ist, dass damit nur bekannte Schwachstellen simuliert werden können. Sogenannte *Zero-Day* Angriffe, d.h. Angriffe auf bisher unbekannte Schwachstellen, können mit Nepenthes nicht erkannt werden. Der zweite in InMAS integrierte Sensor, siehe Abschnitt 2.2, erlaubt es dem System auch Zero-Day Angriffe zu verarbeiten. Als dritter Sensorer kommen sogenannte *Spamtraps* und *Honeyclients* zum Einsatz. Diese Sensoren zielen auf neuartige Angriffsvektoren ab, bei denen Client-Anwendungen wie beispielsweise ein Webbrowser angegriffen werden. Eine Beschreibung dieser Sensoren erfolgt in Abschnitt 2.3.

2.1 Nepentes Erweiterungen

Das Nepenthes-System existiert zur Zeit als einsatzfähiger Prototyp und wird von zahlreichen Organisationen eingesetzt (Bächer, Freiling, Holz, Dornseif, & Kötter, 2006). Im Rahmen des InMAS-Projekts wurde Nepenthes um zahlreiche Module

erweitert, um mehr Schwachstellen zu unterstützten und eine Kopplung mit weiteren Systemen wie beispielsweise dem IAS zu ermöglichen.

Eine zweite zentrale Erweiterung stellt die Integration von Anonymisierungs- und Pseudonymisierungstechniken dar. Nepenthes sammelt standardmäßig Daten, ohne datenschutzrechtliche Aspekte in Betracht zu ziehen. Erst durch die Erweiterung des Systems ist es möglich, dass Nepenthes Daten in datenschutzkonformer Weise sammelt und ausgibt/weitergibt. Dazu werden die gesammelten Informationen entweder pseudonymisert, anonymisiert oder sensitive Informationen werden komplett gelöscht. So können auch Firmen ohne rechtliche Bedenken einen Honeypot-Sensor innerhalb ihres Firmennetzes aufstellen, intern alle Daten betrachten und analysieren und extern nur unbedenkliche Daten weitergeben. Auf diese Weise soll die Hemmschwelle bei Firmen gesenkt werden, einen Honeypot-Sensor in ihr Firmennetzwerk zu integrieren.

Das Aufstellen eines neuen InMAS-Sensors sollte so einfach wie möglich sein. Ideal wäre eine vorkonfigurierte Hardware oder eine Live-CD, die ohne Administrationsaufwand in ein lokales Netz integriert werden kann. Dies ist ein wichtiges Kriterium, um die Akzeptanz bei Firmen zu steigern und Berührungsängste abzubauen. Eine Live-CD mit einem vorkonfigurierten InMAS-Sensor wurde im Rahmen des Projekts bereits implementiert (*Nepox*). Diese CD ermöglicht das Aufsetzen eines neuen Sensorknotens innerhalb weniger Minuten und verringert dadurch auch die Kosten, da zum Aufsetzen und Betrieb eines Sensors nur wenig Aufwand nötig ist.

2.2 Zero-Day Erkennung von Malware

Nepenthes stellt sogenannte *Schwachstellen-Module* zur Verfügung. Diese emulieren Schwachstellen, die von Malware ausgenutzt werden können. Für neu entdeckte Schwachstellen wurden solche Module bisher händisch gefertigt. Durch die Integration von Projekten wie Argos (Portokalidis, Slowinska, & Bos, 2006), Vigilante (Costa, et al., 2005), ScriptGen (Leita, Dacier, & Massicotte, 2006) und die Verwendung von High-Interaction-Honeypots ist es jedoch möglich, Schwachstellen-Module automatisch zu generieren. Entsprechende Vorarbeiten am Institut Eurecom (Frankreich) implementieren solch einen Ansatz und ermöglichen damit, *automatisch* und ohne menschliche Interaktion neue Schwachstellen in Netzwerkdiensten zu detektieren. Im Rahmen eines Frühwarnsystems kann so ein Mechanismus geschaffen werden, um frühzeitig auf neue Bedrohungen reagieren zu können. InMAS wurde um eine auf den Vorarbeiten des Instituts Eurecom basierende Sensorik zur Erkennung von Zero-Day Angriffen erweitert. Um die Fähigkeit des Erkennens von neuartigen Angriffen noch weiter auszubauen, wurde die Sensorik von InMAS um eine weitere Honeypot-Lösung mit dem Namen *Honeybow* erweitert. Honeybow nutzt keine Emulation wie Nepenthes, sondern ein natives System zusammen mit spezieller Software zur Überwachung des Systemzustands. Damit ist Honeybow in der Lage, auch neuartige Angriffe zu erkennen, da keine Signaturen von Angriffen benötigt werden.

2.3 Spamtraps und Honeyclients

Neben den bereits erwähnten Malwarearten, die sich über Schwachstellen in Netzwerkdiensten verbreiten, existieren weitere Verbreitungsvektoren. Beispielsweise nutzen *Storm Worm* (Holz, Steiner, Dahl, Biersack, & Freiling, 2008) und andere Arten von Malware E-Mail zur Weiterverbreitung. Dabei wird die Malware entweder als Dateianhang verschickt oder in der E-Mail Nachricht ist ein Verweis enthalten, der zur Malware führt. Eine weiterer Verbreitungsvektor sind bösartig präparierte Webseiten, die eine Schwachstelle im Browser des Besuchers ausnutzen und dadurch Kontrolle über dessen Rechner erlangen (sogenannter *Drive-by Download*).

Des Weiteren werden auch gezielt Schwachstellen in Client-Applikationen wie zum Beispiel dem Adobe Reader oder Microsoft Word für Angriffen ausgenutzt: Der Angreifer sendet zielgerichtet eine E-Mail mit einem bösartigen Dateianhang, der eine Schwachstelle in einer entsprechenden Client-Applikation ausnutzt, an sein Opfer. Auch für diesen Verbreitungsweg von Malware verfügt InMAS über eine spezielle Sensorik, um mehr Informationen über diese Art der Malware-Verbreitung zu sammeln. Hierzu werden sogenannte *Spamtraps* angelegt: Eine Spamtrap ist eine Art Honeypot, der nur E-Mail Spam-Nachrichten anziehen soll. Da diese E-Mail-Postfächer nicht für eine reguläre Kommunikation genutzt werden, ist jede eintreffende E-Mail als Spam zu betrachten. Hierbei kann neben einzelnen E-Mail-Postfächern auch eine komplette Domain registriert und als Spamtrap eingesetzt werden. Alle auf den verschiedenen Spamtraps eintreffenden E-Mails werden automatisch analysiert: Verdächtige Dateianhänge werden mit der CWSandbox analysiert und in der Nachricht enthaltene URLs automatisch mit Honeypots untersucht. Zur Analyse von bösartigen Webseiten werden *Honeyclients* benutzt. Dies ist eine Honeypot-Art, der automatisiert Webseiten untersucht und Informationen darüber sammelt, ob die besuchte Webseite versucht den Browser zu kompromittieren. Dazu kann entweder ein *Low-Interaction* oder ein *High-Interaction* Ansatz verfolgt werden:

- Low-Interaction Honeyclients sind *Crawler*, die das Internet systematisch nach verdächtigen Webseiten absuchen. Die gefundenen Webseiten werden mit verschiedenen Tools untersucht und bösartige Webseiten automatisch klassifiziert.
- High-Interaction Honeyclients untersuchen Webseiten mit einem instrumentierten Browser und surfen diese automatisiert an. Durch eine Integritätsprüfung wird festgestellt, ob eine gegebene Webseite eine Schwachstelle im Browser ausgenutzt hat und welche Veränderungen am System durchgeführt wurden.

Im Rahmen des InMAS-Projekts wurden diese beiden Ansätze evaluiert und ein entsprechender Honeyclient in die Sensorik integriert. Mit dieser Erweiterung ist InMAS in der Lage, neben Bots und Würmern auch Informationen über andere Arten von Malware zu sammeln. Dies ist besonders relevant da die Angriffe auf Client-

Applikationen in den letzten Jahren deutlich zugenommen haben und eine immer größere Gefahr darstellen.

3. Analyse und Auswerungssystem (*Malware Analysis*)

Das Auswerungssystem von InMAS kombiniert eine Vielzahl von Analysewerkzeugen, um die gesammelten Informationen in eine für einen Analysten verständliche Form zu bringen. Dabei ist es wichtig, dass die mit Hilfe von Honeypots gesammelten Kopien von Schadsoftware zu einem hohen Grad *automatisiert* analysiert werden, da die Notwendigkeit von menschlicher Interaktion im Rahmen eines Frühwarnsystems möglichst gering gehalten werden sollte. Neben statischen Analysenwerkzeugen wie Antivirenscannern stellt die CWSandbox, ein Tool zur dynamschen Verhaltensanalyse von Malware, das zentrale Werkzeug von InMAS dar. Diese Softwarelösung liefert für jede Malware einen detailierten Verhaltensreport, aus dem ein Vielzahl an Informationen gewonnen werden kann. Auf die speziellen Anforderungen eines Frühwarnsystems an die dynamische Analyese und die in InMAS integrierten Erweiterungen, wie beispielsweise die Benutzersimulation, wird in Abschnitt 3.1 eingegangen.

Die Schnittstelle zwischen den Analysten und dem Frühwarnsystem bildet das *Statistik-Backend*, siehe Abschnitt 3.2. Es bereitet die Ergebnisse der verschiedenen Analysewerkzeuge auf und stellt diese in übersichtlicher Form dar. Zusätzlich werden dem Analysten die Ergebnisse veschiedener Metriken präsentiert, die ihn bei der Gefahrenabschätzung und Erkennung von Anomalien unterstützen.

3.1 Dynamische Analyse mittels CWSandbox

Die bereits existierende Software CWSandbox ermöglicht es, eine automatisierte, verhaltensbasierte Analyse von Malware durchzuführen. Dazu wird die Schadsoftware in einer kontrollierten Umgebung ausgeführt und zur Laufzeit beobachtet, welche Aktionen ausgeführt werden. Das Resultat eines solchen Analyselaufs ist ein für Menschen verständlicher Report der beobachteten *Verhaltensweise* der Malware.

Zum Einsatz in einem Frühwarnsystem muss ein spezieller Fokus auf die Netzwerkkommunikation von Malware gelegt werden. Möglichst schnell und detailliert muss herausgefunden werden, wie die Malware mit anderen Rechnern kommuniziert und welche Arten von Daten dabei übertragen werden. Dies ermöglicht es auch, Kommunikationsprofile in unterschiedlichen Phasen der Malware-Aktivität (Verteilung/Verbreitung, Nachladen bzw. Ausführung der Schadroutine) zu erstellen. Nur so kann das Gefährdungspotenzial realistisch eingeschätzt werden und eventuelle Gegenmaßnahmen getroffen werden. Diese Art von Informationen kann beispielsweise vom Internet-Analyse-System genutzt werden, um Angriffe zu erkennen. Des Weiteren helfen detaillierte Ergebnisse über die Veränderungen des infizierten Systems, schnell Signaturen zur Malware-Erkennung erstellen zu können. All diese Arten von Informationen müssen auf verschiedenen Ebenen innerhalb des Betriebssystems gesammelt und analysiert werden.

Nicht immer ist es wünschenswert, dass die zu analysierende Malware-Datei auf das Internet zugreift. Dies kann zum Beispiel der Fall sein, wenn die zu untersuchende Schadsoftware auf keinen Fall Informationen nach außen senden darf. Damit die Malware dennoch mit dem Internet interagieren kann, ist eine *Emulation* des Netzes nötig. Dies kann durch die Emulation eines DNS- oder HTTP-Servers geschehen. Weiter kann es wünschenswert sein, dass die Kommunikation der Malware nach außen über einen Proxy läuft, um dadurch eine Filterung der übertragenen Daten durchzuführen. Die Kombination dieser beiden Aspekte wird innerhalb von InMAS in einem Modul realisiert (*TrumanBox*). Dieses Modul ist in der Lage, ein komplexes Netz zu emulieren (inklusive Dienste wie DNS, IRC oder SMTP) und eine Filterfunktion zur Verfügung zu stellen.

Einige Arten von Malware führen erst dann ihre Schadfunktion aus, wenn bestimmte Bedingungen erfüllt sind − diese Bedingungen werden *Triggerbedingungen* genannt. Neben einer IP-Adresse aus einer bestimmten IP-Range oder der Aktivierung der schadhaften Funktion ab einem bestimmten Datum, können diese Bedingungen auch vom Verhalten des Systembenutzers abhängen, wie zum Beispiel dem Besuch einer Online-Banking Webseite, der Kommunikation mittels Instant Messenger Programmen oder dem Starten eines E-Mail Programms. Um möglichst viel Malware zur Ausführung ihrer Schadfunktion zu bringen, muss das Benutzerverhalten simuliert werden. Dazu ist es nötig, ein Modul zu entwickeln, das einen Benutzer des Systems simuliert und typische Aktionen ausführt. Ein funktionierender Prototyp eines solchen Moduls zur Benutzersimulation wurde im Rahmen des Projekts implementiert (*SimUser*). Das Modul kann auch unabhängig von der CWSandbox betrieben werden und simuliert typische Verhaltensmuster eines Benutzers wie beispielsweise das Ansurfen einer Webseite oder das Versenden einer E-Mail Nachricht.

3.2 Statistik-Backend

InMAS sammelt umfangreiche Daten über verschiedene Arten von Malware. Dies beinhaltet zum Beispiel die IP-Adresse des Angreifer oder den angegriffenen Netzwerkdienst. Diese Daten werden zusätzlich durch die mittels CWSandbox generierten Analysen und andere Datenquellen angereichert. Insgesamt wird eine große Menge an Daten gesammelt, die für die Erstellung eines Lagebilds wichtig sind. Beispielsweise wurden im Jahr 2008 durch InMAS insgesamt mehr als 500GB an Analysedaten erzeugt. Durch eine statistische Analyse dieser Daten können Veränderungen der Nutzung bestimmter Netzwerkdienste oder das Auftreten von Netzwerkanomalien frühzeitig erkannt werden. Diese Informationen sind wichtig für das Lagezentrum, können aber auch zur Erstellung von Signaturen (auf Netzwerkebene oder auch auf Systemebene) genutzt werden.

Im Rahmen dieses Teilprojekts wird ein Backend für das *Malware Analysis Repository* entwickelt. Dieses erlaubt den Analysten, die Analyseergebnisse der verschiednen Werkzeuge mit Hilfe einer zentralen Schnittstelle abzufragen und liefert somit einen Überblick über das gesammte Frühwarnsystem. Die graphische Aufbereitung der Ergebnisse lenkt die Aufmerksamkeit der Analysen dabei auf die zentralen Ereignisse

und hilft diese schnell zu erkennen. Gleichzeitig liefert das Statistik Backend auch detailierte Informationen zu den aufgezeichneten Angriffen und den auf der gesammelten Malware durchgeführten Analysen. Des Weiteren werden momentan verschiedene Techniken aus dem Bereich des Maschinellen Lernens – vor allem die Aspekte Clustering und Klassifizierung – in das Statistik-Backend integriert. Dies erlaubt es, aus der großen Datenmenge weitere Informationen zu gewinnen und ermöglicht, eine weitestgehend automatisierte Auswertung der gesammelten Informationen.

Das Statistik-Backend wird in Abschnitt **Error! Reference source not found.** im Rahmen der Projektergebnisse noch weiter erläutert beziehungsweise in Abbildungen dargestellt.

3.3 Verlässlichkeitsmetriken

Messungen eines Frühwarnsystems haben zur Zeit nur eine begrenzte Aussagekraft für den allgemeinen Zustand des deutschen Internet. Die grundlegende Fragestellung dieses Teilaspekts des Projekts lautet: *Wie kann man die Repräsentativität/Aussagekraft der Messungen beziffern?*

Die einzig bekannte Maßzahl einer repräsentativen Abdeckung ergab sich bisher aus einer durch die Autoren betreuten Diplomarbeit an der RWTH Aachen und bezieht sich auf den Einsatz von Nepenthes als Intrusion Detection Sensor: Demnach ist es ausreichend, Nepenthes-Sensoren in einem Netzwerk auf 0,1% der vorhandenen IP-Adressen zu verteilen, um dieselbe Menge an Vorfällen zu registrieren, die man auch durch ein Filtern des kompletten Netzverkehrs am zentralen Gateway erkannt hätte. Eine ähnliche Fragestellung wird im Kontext dieses Projektes für deutlich größere Netze untersucht werden.

4. Ergebnisse

In diesem Abschnitt werden beispielhaft einige Ergebnisse, die aus dem realen Betrieb von InMAS gewonnen wurden, dargestellt. Momentan befindet sich das Komplett-System noch in Entwicklung, die Fertigstellung ist für Dezember 2009 geplant. Dieser Abschnitt gibt einen Überblick über den Systemzustand im Januar 2009 und zeigt exemplarisch die einzelnen Aspekte des Systems.

Die Benutzerschnittstellen von InMAS wurden einheitlich als Web-Schnittstellen realisiert, um einen einfachen Zugriff von verschiedenen Standorten aus zu ermöglichen. Dabei wurde darauf geachtet, dass alle Schnittstellen ähnlich aufgebaut sind und dem Prinzip von Ben Shneiderman folgen: „Overview first, zoom and filter, then details-on-demand.“ Ein Analyst erhält also auf der Anfangsseite zunächst einen Überblick über den aktuellen Systemzustand und die von der Sensorik zuletzt gesammelten Informationen. Eine Filterung nach bestimmten Aspekten (beispielsweise IP-Adressen oder Angriffsversuche) ist von der Übersichtsseite aus möglich und auch die detaillierte Betrachtung eines Angriffs wird unterstützt. Ein Analyst kann also flexibel auf die gesammelten Informationen zugreifen und bei

Bedarf auch einzelne Datensätze anschauen. Abbildung 2 zeigt exemplarisch den Aufbau der Schnittstelle zum *Raw Malware Repository* in der Detail-Ansicht: Zu einem bestimmten Angriff werden alle verfügbaren Informationen wie beispielsweise die Art des Angriffs oder Meta-Informationen zur IP-Adresse des Angreifers angezeigt. Desweiteren wird eine Korrelation mit anderen Angriffen durchgeführt, um festzustellen welche weiteren Angriffe noch von der IP-Adresse des Angreifers aus beobachtet werden konnten.

Abbildung 3 zeigt den Aufbau der Schnittstelle zum *Spamtrap Repository*. Ein Vergleich mit Abbildung 2 zeigt den sehr ähnlichen Aufbau beider Systeme, wodurch sich die Vorgehensweisen bei verschiedenen Analysen so homogen wie möglich gestalten. In diesem Beispiel sieht man das *Analysetool*, das einige vorkonfigurierte Abfragen zur Verfügung stellt. Mit diesem Tool können statistische und zeitliche Auswertungen über die gesammelten Informationen durchgeführt werden, um spezielle Vorfälle genauer analysieren zu können. Insgesamt wurden etwa 25 solche Abfragen implementiert, um typische Abfragen eines Analysten direkt umsetzen zu können.

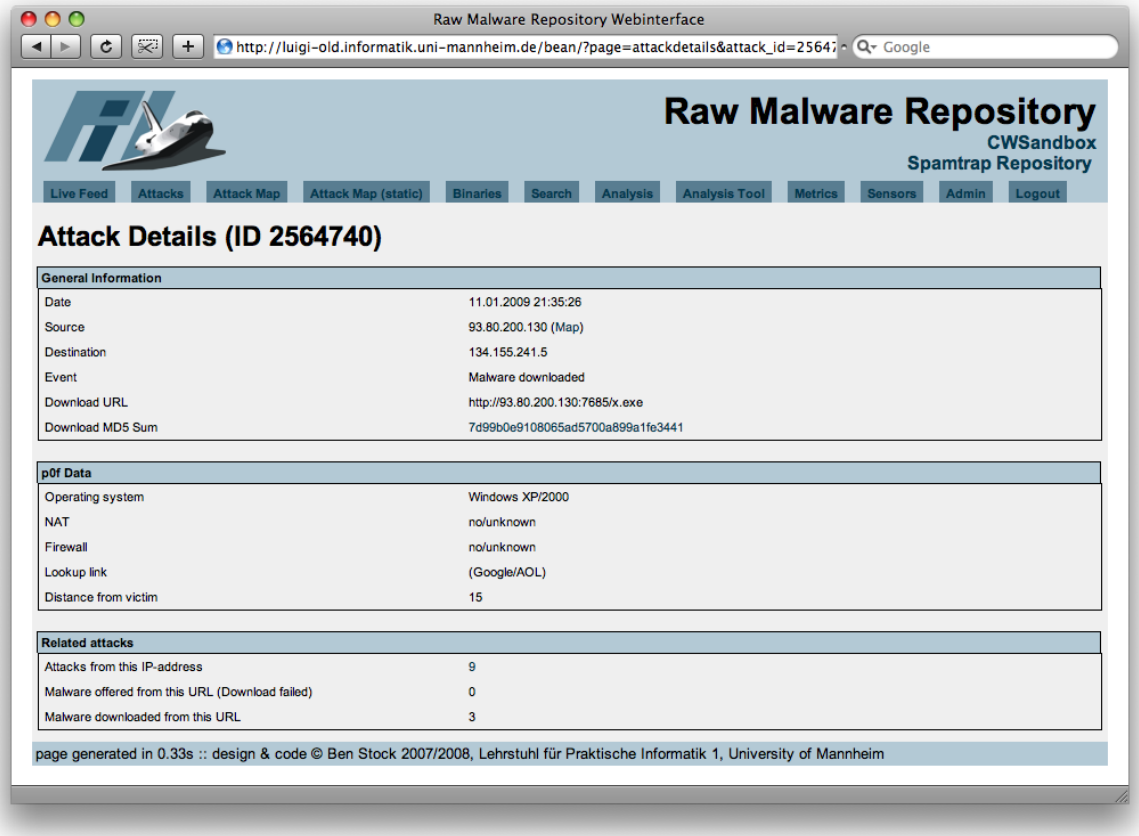

Abbildung 2: Analysten-Sicht des Raw Malware Repository

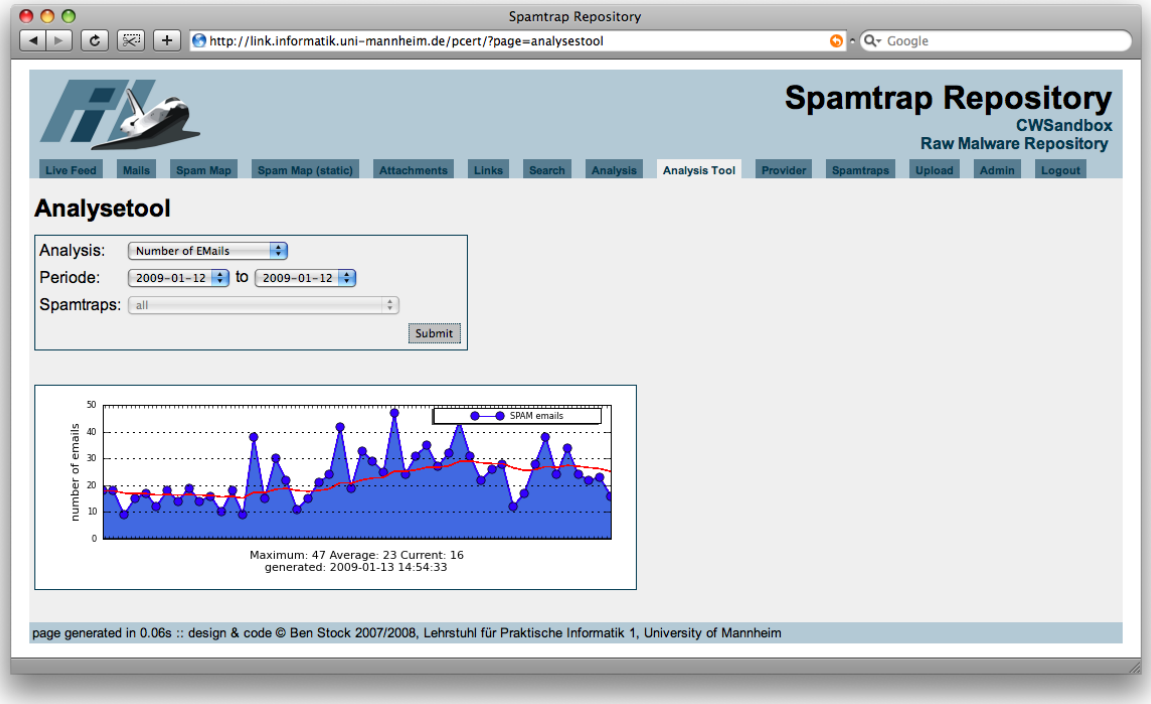

Abbildung 3: Analysten-Sicht für das Spamtrap Repository

Zur Visualisierung der gesammelten Informationen werden verschiedene Techniken wie beispielsweise Kreis- oder Balkendiagramme, Zeitverläufe oder die geographische Aufbereitung der Daten mittels *Google Maps* benutzt. Dies ermöglicht einem Analysten den schnellen Überblick über die Sensorik und erlaubt das frühzeitige Erkennen neuer Trends und Anomalien innerhalb des Netzes. Beispielhaft wird im Folgenden die geographische Aufbereitung von IP-Adressen vorgestellt, andere Visualisierungstechniken arbeiten prinzipiell ähnlich. IP-Adressen werden von InMAS häufig beobachtet, beispielsweise durch Exploit-Versuche gegen die Honeypots oder Header-Daten aus Spam-Nachrichten. Für alle diese IP-Adressen wird zunächst ein *GeoIP-Lookup* durchgeführt, d.h. mit Hilfe spezieller Datenbanken wird die Latitude und Longitude zu einer gegebenen IP-Adresse bestimmt. Diese Informationen werden dann auf eine Weltkarte projeziert, um einen Überblick über die Standorte der Angreifer zu bekommen. Abbildung 4 zeigt exemplarisch die geographischen Standorte von Maschinen, die zur InMAS-Sensorik Spam-Nachrichten gesendet haben. Über die Farbe des Markers wird dabei kodiert, wieviele Spam-Nachrichten von einem bestimmten Standort empfangen wurden.

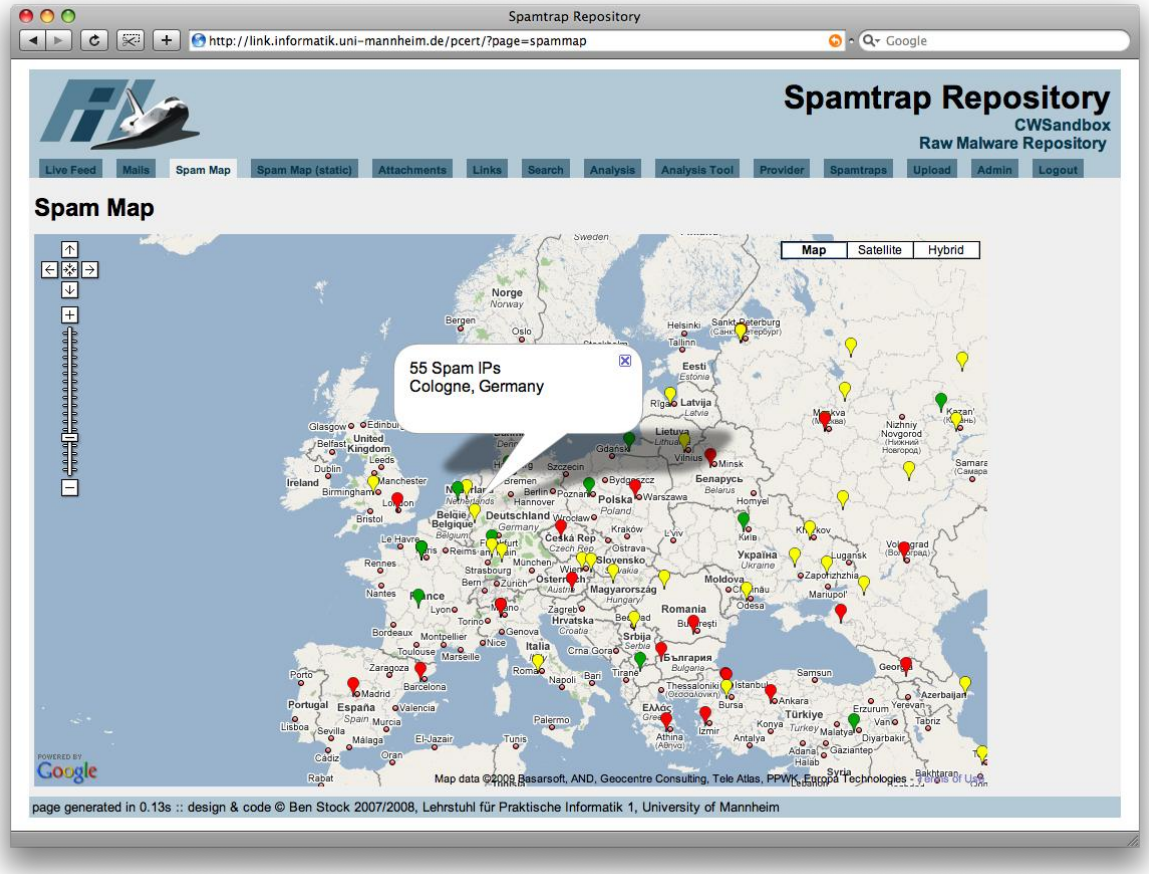

Abbildung 4: Geographische Aufbereitung der Informationen im Spamtrap Repository

Um schnell auf die gesammelten Informationen zugreifen zu können werden die gesammelten Daten in einem *Data Warehouse* abgespeichert. Ein Data Warehouse ist ein spezielle Form einer Datenbank, deren Design für flexible Anfragemöglichkeiten und gute Performance bei hohen Datenmengen optimiert ist. Zentrale Komponente zur Umsetzung einer solchen Datenbank ist der sogenannte *Data-Cube*. Über dieses Konzept des Datenwürfels wird es ermöglicht, große Datenmengen schnell und intuitiv multidimensional zu analysieren. Kernpunkt hierbei ist die Aufteilung der Daten in Fakten- und Dimensionstabellen, die in einem Sternschema angeordnet sind. Diese Bezeichnung rührt daher, dass in der grafischen Darstellung Dimensionstabellen sternförmig um die Faktentabellen angeordnet sind. Insgesamt wurden 24 Dimensionstabellen erzeugt, die beispielsweise Zugriffe auf das Dateisystem oder das Netzwerk sowie Veränderungen an der Windows Registry abspeichern.

Im Oktober 2008 wurden mittels CWSandbox insgesamt 46.622 Analysereports für die durch InMAS gesammelten Dateien erzeugt. Diese Reports erfordern 4,89 GB Speicherplatz. Zum effizienten Zugriff auf die Analysereport wurden diese in die Kern-Informationen zerlegt und dann in das Data Warehouse eingefügt. Die so erzeugte Datenbank für alle Analysen enthält 8.495.783 Einträge und hat eine Speichergröße von 780 MB.

Das Data Warehouse ermöglicht dann die schnelle Abfrage von bestimmten Informationen. Beispielsweise kann eine Anfrage gestellt werden, wie oft in einem bestimmten Zeitraum die analysierten Schadprogramme auf eine bestimmte Datei zugriffen haben (siehe Abbildung 5). Statistische Abfragen können mit Hilfe des Data Cubes schnell durchgeführt werden. Beispielsweise benötigt die Berechnung der zehn am häufigsten gelöschten Dateien im Oktober 2008 nur 6,4 Sekunden.

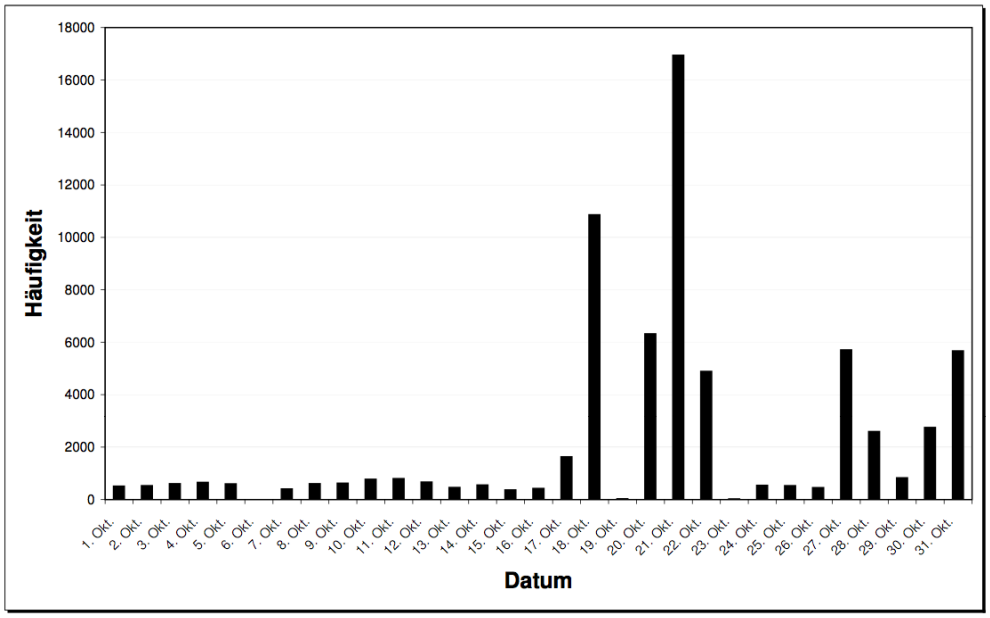

Abbildung 5: Verwendung der Datei "autoexec.bat" im Oktober 2008

Eine zentrale Frage bei der Einschätzung von Gefährdung in einem Frühwarnsystem ist, wie diese überhaupt gemessen werden kann. Es müssen *Kennzahlen* identifiziert werden, nach denen bewertet werden kann, wie hoch das Level der Gefährdung ist. Im Rahmen des InMAS Projekts wurden einige Kennzahlen und *Sicherheitsmetriken* ausgearbeitet, um einzelne Dimensionen der Gefährdung durch automatisierte Malware einschätzen zu können. Aus diesen Einzelkennzahlen wird eine Gesamtmetrik bestimmt, die eine allgemeine Aussage über das Gefährdungslevel macht. Alle entwickelten Kennzahlen sind mit dem Ziel entwickelt worden, Möglichkeiten aufzuzeigen, wie eine Beurteilung des Gefährdungslevels anhand einzelner Kennzahlen erfolgen kann.

Ein wichtiger Aspekt bei der Entwicklung der Metriken war – wie bereits bei den anderen Komponenten des Projekts – die Automatisierbarkeit. Alle Metriken können automatisiert mit den durch InMAS gesammelten Daten bestimmt und mittels der Web-Schnittstelle abgerufen werden. Dies kann, ebenso wie bei den Ergebnissen der einzelnen Datenbanken sowohl mit den jeweils aktuellen, als auch mit historischen Daten erfolgen. Für die Berechnung der Metriken wird jeweils der Zeitraum der letzten sieben Tage zugrunde gelegt. Dadurch wird eine Vergleichbarkeit einzelner Ausprägungen einer Metrik hergestellt. Ein kleinerer Zeitraum würde unter Umständen zu großen Schwankungen in den Ausprägungen der Metrik führen. Wäre der Zeitraum größer gewählt, beispielsweise 30 Tage, würden viele historische Daten berücksichtigt, die nicht die aktuelle Sicherheitslage widerspiegeln. Gerade im

schnelllebigen Bereich der IT-Sicherheit und für den Einsatz in einem Frühwarnsystem ist die Verwendung aktueller Daten von großer Bedeutung.

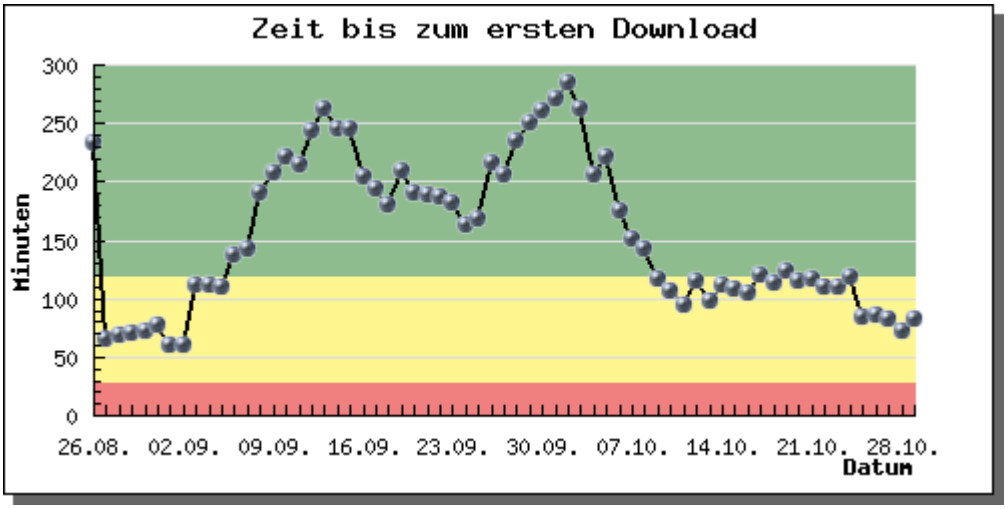

Abbildung 6: Kennzahl zur "Zeit bis zum ersten Download" im zeitlichen Verlauf

Abbildung 6 zeigt ein Beispiel für eine von InMAS benutzte Metrik: Die Berechnung dieser Metrik erfolgt, indem zunächst zu jeder neuen Sensoren-IP-Adresse der letzten sieben Tage die Zeit in Minuten bis zum nächstfolgenden Download eines Malware Binaries bestimmt wird. Der Wert der Metrik ist der Durchschnitt aus diesen Einzelwerten. Angegeben wird die Metrik in Minuten. Es wird also über alle Sensoren gemittelt dargestellt, wie lange es durchschnittlich gedauert hat, bis ein Sensor erfolgreich eine Schadsoftware heruntergeladen hat. Dies ist ein Indiz für die Scanning-Aktivität im Internet. Je niedriger der Wert für die durchschnittliche Dauer bis zum herunterladen einer Schadsoftware, um so höher die Scanning-Aktivität und dementsprechend auch das aktuelle Gefährdungslevel.

Aktuell werden in InMAS sechs Kennzahlen berechnet, die die aktuelle Gefahrenlage im Internet wieder spiegeln. Neben der bereits erwähnten "Zeit bis zum ersten Download" sind dies:

- der Anteil von veralteten Betriebssystemen
- die Anzahl unterschiedlicher Angreifer
- die Anzahl verschiedener Binaries
- die Anzahl unbekannter Binaries
- den Anteil unerkannter neuer Binaries

Zur Einschätzung des aktuellen Gefährdungslevels durch automatisierte Malware im Internet wird in InMAS eine auf diesen sechs Kennzahlen basierende Gesamtmetrik berechnet. Da die sechs Kennzahlen in ihrer Art sehr verschieden sind - sie bilden sehr unterschiedliche Bereiche der Sicherheitsbetrachtung ab und haben verschiedene, zum Teil unvereinbare Maßeinheiten - können nur die Interpretationen der einzelnen Metriken verknüpft werden.

Die Bestimmung der Gesamtmetrik erfolgt über die Ausprägungen der Einzelmetriken gemessen in Kategorien ihrer Interpretation. Jede dieser Einzelmetriken betrachtet einen spezifischen Bereich der Sicherheitslage im Internet in Bezug auf automatisierte Malware. Daher kann keine Priorisierung unter den einzelnen Kennzahlen vorgenommen werden. Sobald eine der Metriken einen hohen Wert annimmt, ist eine Dimensionen der Gefährdung hoch. Um dieser Tatsache gerecht zu werden, wird die Gesamtgefährdung bestimmt als das jeweilige Maximum aller Kategorienwerte der Einzelmetriken. Ist also eine der Kennzahlen im roten Bereich, ist auch die Gesamtmetrik hoch. Nur, wenn in allen Dimensionen ein niedriges Gefährdungslevel vorliegt, kann auch die Gesamtgefährdung als niedrig eingestuft werden.

5. Zusammenfassung und Ausblick

Resultat dieses Forschungsprojekts ist eine weltweit einmalige Sammlung an Analyse- und Statistikwerkzeugen für sich automatisiert verbreitende, bösartige Software. Die Analyseergebnisse befähigen das nationale Frühwarnsystem weitaus besser die Gefahren zu erkennen, zu klassifizieren und einzuschätzen, die durch automatisiert verbreitende Malware entstehen. Die gewonnen Daten können zur weiteren Analyse von aktuellen Entwicklungen herangezogen werden und Fragen nach dem Ursprung eines Wurmbefalls, nach den typischen geographischen Ausbreitungspfaden sowie nach dem zeitlichen Verlauf einer Ausbreitung beantworten.

InMAS wird kontinuierlich weiterentwickelt. Hierbei wird versucht, aktuelle Gefahren zu erkennen und das Fühwarnsystem entsprechend anzupassen. Im Bereich des Sensorsystems sei hierbei auf die Integration der Honeyclients und der Zero-Day Erkennung verwiesen. Auch das Analyse- und Auswertungsystem wird regelmäßig erweitert. Aktuell wird eine generische Packeranalyse in InMAS integriert: Moderne Malware ist häufig *gepackt*, das heisst, dass eigentliche Binary ist komprimiert und entpackt sich erst zur Laufzeit. Eine andere Möglichkeit des Verschleierns von Binaries sind *Crypter*, die das Binary verschlüsseln und zur Laufzeit entschlüsseln. Beide Ansätze ermöglichen es einem Angreifer, signaturbasierte Antivirenlösungen zu umgehen. Um im Rahmen eines Frühwarnsystems solche Angriffe schnell und automatisiert zu erkennen, muss ein Ansatz geschaffen werden, um die Verwendung von Packern oder Cryptern gezielt zu entdecken und das Binary möglichst vollständig zu entpacken beziehunsweise entschlüsseln. Für das Gesamtprojekt ist eine solche Komponente nötig, um im *Malware Analysis Repository* auch eine entpackte/entschlüsselte Version der Malware ablegen zu können. Nur mit einer solchen Version der Malware kann ein Analyst einfach eine tiefergehende manuelle Analyse durchführen und das Gefährdungspotenzial abschätzen. Desweiteren erlaubt ein generischer Entpacker, auch Querverweise zwischen einzelnen Binaries zu entdecken und so weitere mögliche Angriffspfade zu erkennen.

Literaturverzeichnis

Bächer, P., Freiling, F., Holz, T., Dornseif, M., & Kötter, M. (2006). The Nepenthes Platform: An Efficient Approach to Collect Malware. *Proceedings of 9th Symposium on Recent Advances in Intrusion Detection (RAID'06).*

Costa, M., Crowcroft, J., Castro, M., Rowstron, A., Zhou, L., Zhang, L., et al. (2005). Vigilante: End-to-end containment of Internet worms. *SIGOPS Oper. Syst. Rev.*

Holz, T., Steiner, M., Dahl, F., Biersack, E., & Freiling, F. (2008). Measurements and Mitigation of Peer-to-Peer-based Botnets: A Case Study on Storm Worm. *USENIX LEET '08.*

Leita, C., Dacier, M., & Massicotte, F. (2006). Automatic Handling of Protocol Dependencies and Reaction to 0-Day Attacks with ScriptGen Based Honeypots. *Recent Advances in Intrusion Detection, 9th International Symposium, RAID 2006.*

Moore, D., Shannon, C., & Claffy, K. (2002). Code-Red: a case study on the spread and victims of an internet worm. *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment.*

Portokalidis, G., Slowinska, A., & Bos, H. (2006). Argos: An Emulator for Fingerprinting Zero-Day Attacks. *Proc. ACM SIGOPS EUROSYS'2006.*

Willems, C., Freiling, F., & Holz, T. (2007). CWSandbox: Towards Automated Dynamic Binary Analysis. *IEEE Security and Privacy.*