Forschungsprojekt
Frühwarnsystem für Botnetze

Dr. Alexander K. Seewald, Doz.Dr. Wilfried N. Gansterer
05.02.2008

1 Kurzbeschreibung des Projekts

Dieses Projekt geht auf die Initiative von Dr. Alexander Seewald zurück und wird in Kooperation mit der Arbeitsgruppe von Dr. Wilfried Gansterer am Research Lab Computational Technologies and Applications der Universität Wien durchgeführt. Sowohl Dr. Gansterer als auch Dr. Seewald weisen mehrjährige Forschungserfahrung im Bereich Anti-Spam auf.

Existierende Forschungsinitiativen konzentrieren sich hauptsächlich auf die Erkennung und Abwehr von unerwünschten oder potentiell schädlichen E-Mail Nachrichten (hier vereinfachend mit dem Begriff *Spam* zusammengefasst), wie z.B. auch das derzeit laufende IPA-Netidee-Projekt *Anti-Phishing* von Dr. Wilfried Gansterer. In diesem Projekt soll als wichtige Ergänzung dazu ein wichtiger anderer Aspekt, nämlich die (vorbeugende) Identifizierung und Früherkennung des Ursprungs von Spam eingehend untersucht werden.

Der Großteil des Spam wird derzeit von großen Netzen gekaperter Rechner unschuldiger Benutzer, die über die Ausnutzung von Sicherheitslücken mit entsprechender Malware infiziert wurden, der Kontrolle von Spammern unterstehen, und in sogenannten Bot-Netzen organisiert sind, gesendet. Dieselben Netze werden mittlerweile auch für verschiedene andere Attacken, wie z.B. Angriffe auf die Netzwerkinfrastruktur und auf den Spammern missliebige Unternehmen (z.B. http://de.wikipedia.org/wiki/Blue_Frog) verwendet.

Es ist leider noch immer viel zu wenig über die Lebenszyklen, den Aufbau, die Verwendung und den Abbau von Bot-Netzen bekannt, um geeignete Schutz- und Gegenmaßnahmen zu ergreifen. Ziel dieses Projektes ist es, basierend auf existierender Expertise sowie unter Verwendung existierender Infrastruktur der Universität Wien als auch neuer leistungsfähiger Hardware, Tools für eine umfassende passive Analyse von Bot-Netz-Traffic zu konzipieren und zu entwickeln.

Diese Tools sollen eine vollständig automatisierte Unterstützung für die bestmögliche Identifikation und das laufende Tracking von Bot-Netzen und deren Verwendung bereitstellen. Diese automatisierte Lösung wird anschließend per Open Source und Webformular der Allgemeinheit zur Verfügung gestellt und ermöglicht die Entwicklung darauf aufbauender Anwendungen, wie z.B. eines Frühwarnsystems für die Besitzer von gekaperten Rechnern bzw. für österreichische Internet Service Provider, eines Systems für das Bereitstellen wertvoller Informationen für die österreichische Justiz zur Verfolgung von Spammern und Betreibern von Bot-Netzen, sowie umfassender und aktueller Blacklists zum Aussortieren von Spam dieser Bot-Netze.

2 Projektplan

Die Deadline für Deliverables dieses Projektes wird mit dem Ende der entsprechenden Projektphase plus zwei Wochen festgelegt. Im folgenden werden vier Projektphasen mit entsprechenden Deliverables und Meilensteinen definiert.

2.1 State-of-the-Art passiver Analysewerkzeuge, Installation

Dauer: 3 Monate

Beschreibung: Setup eines Testsystems, umfassende Erhebung des state-of-the-art passiver Analysewerkzeuge wie beispielsweise Siphon, p0f, Fl0p und ethereal; Installation und Konfiguration derselben auf dem Testsystem; Glue-Code Programmierung und Aufbau einheitlicher Interfaces. Aufbau & Test der Netzwerk-Infrastruktur (inkl. Domains & multiple IPs) und Beginn der Archivierung des gesamten Netzwerkverkehrs am Testsystem.

Meilenstein: Vollständig funktionsfähiges Packet-Capturing und wesentlicher passiver Analysewerkzeuge auf dem Analyserechner.

Internes Deliverable: Umfassende Übersicht gängiger passiver Analysewerkzeuge mit Beschreibung der Funktionalität, zugrundeliegender Lizenz, Link zu Homepage und kurzer Installationsanleitung. Umfang 5-10 Seiten in engl. Sprache, publiziert unter CreativeCommons Attribution Share Alike (by-sa).

2.2 Erste Analysen des Netzwerkverkehrs

Dauer: 6 Monate

Beschreibung: Erste Analysen des archivierten Netzwerkverkehrs mit den Analysewerkzeugen, Vergleich mit bekannten Bot-Netzen und Visualisierung der Aktivitätsmuster. Aufbau von Trainingsdaten für Bot-Erkennung und Malware-Identifikation basierend auf Fingerprints des Traffics bekannter Bot-Netze. Dabei werden die benötigten Schritte zuerst manuell getestet und gegebenenfalls adaptiert, und die Automatisierung soweit wie möglich bereits vorbereitet.

Meilenstein: Korrelationen zwischen Netzwerkverkehr-Features und der Präsenz von gekaperten Rechnern, die Botnets angehören; verschiedene Aktivitätsmuster zur Erkennung von spezifischen Bots und/oder Botnetzen; Hypothesen zur eindeutigen Identifikation von Botnetzen.

2.3 Automatisierung der manuellen Analysen, Zurverfügungstellung über Webformular

Dauer: 2 Monate

Beschreibung: Automatisierung mittels verschiedener Lernsysteme (beispielsweise Support Vector Machines, RIPPER, multinomial Naive Bayes) und/oder Artificial Intelligence Programmiermethoden, Validierung und Evaluierung sowohl mit zuvor archivierten als auch mit neuen Daten.

Meilenstein: Weitgehende Formalisierung und Automatisierung der Ergebnisse aus 2.2 auf einer soliden statistischen Grundlage; vollständig funktionsfähiges autonomes Analysesystem.

Deliverable: Webformular, auf dem sich jeder österreichische Internet-Benutzer über den Status seines eigenen Rechners informieren kann (d.h., ob dieser einem Botnet angehört und – soweit bekannt – weiterführende Informationen über Status & kürzlich erfolgte Verwendung des Botnetzes, assoziertes Malware-Executable, etc.)

2.4 Zusammenstellung einer installierbaren Version

Dauer: 1 Monat

Beschreibung: Zusammenstellung einer installierbaren Version des Gesamtsystems als Open Source (inkl. der automatisierten Analysen aus 2.3), Installationsanleitung und kurze Dokumentation.

Meilenstein: Installierbare Version des Gesamtsystems.

Deliverable: Installierbare Version des Analysesystems im Sourcecode (GPL v3-Lizenz) inkl. Installationsanleitung (1-2 Seiten) und kurzer Dokumentation (ca. 5-10 Seiten, in engl. Sprache). Als Plattform ist Debian Linux Stable (Etch) vorgesehen. Das fertige Analysesystem wird voraussichtlich zumindest die folgenden Funktionen bieten:

- Blacklist von allen getrackten Bots für verbesserte Spam-Erkennung.
- Liste von Botnetzen mit aufgezeichneter Aktivität (Spam, Denial-of-Service, Pump-and-Dump, Phishing etc.), assoziertes Malware-Executable, Größe, historische Zu- und Abgänge.
- Liste von Aktivität außerhalb von Botnetzen (manuelle Portscans, *script kiddies* Angriffe, Spam-Versendung über dedizierte Linux-Rechner, etc.)
- Detailinformationen pro Bot (assoziertes Botnetz und Malware-Executable, Betriebssystem, historische Aktivitäten mit Time-Stamp etc.)

Wissenschaftlich relevante Ergebnisse aus den Phasen 2.2 und 2.3 werden in einschlägigen Konferenzen, Workshops und/oder Journals veröffentlicht.