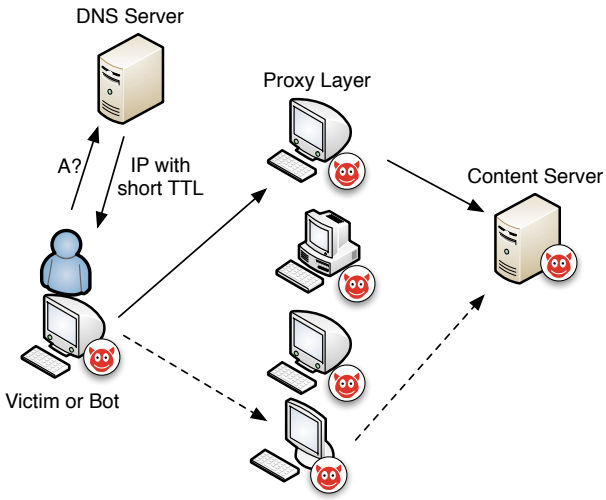
Erkennung von Malware-C&C durch Internet-weite DNS Analyse

Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Hintergrund



Damit mit Malware-infizierte Computer Instruktionen empfangen können, ist ein Command&Control Server (C&C) nötig. Da die Adresse des C&C - eine IP oder Domainname - in jeder Malware enthalten sein muss, ist sie damit natürlich ebenfalls für den Verteidiger sichtbar, und kann geblockt werden oder durch Polizei-aktionen beschlagnahmt werden. In den vergangenen 20 Jahre haben Malware-Autoren daher immer neue Wege gefunden, die Adresse des zentralen Koordinationsknotens zu verschleiern. Feste Domainnamen oder IP Adressen wurden schnell abgelöst durch so genannte Domain Generation Algorithms (DGAs), bei denen Hunderte Domains in kurzen Zeitintervallen algorithmisch generiert werden. Um das Botnetz zu stoppen, müssen daher Verteidiger die jeweils gültigen DGA-Ausgaben vorberechnen und immer wieder aufs neue blockieren. Malware verschleiert jedoch nicht nur die Adresse des C&Cs, sondern auch wo dieser sich gerade befindet. In vielen modernen Botnetzen weist der Domain Name nicht mehr direkt auf den C&C, sondern auf andere Bots, die Anfragen temporär an den eigentlichen Server weiterleiten. Diese so-genannten Proxy Layer verschleiern die Struktur des Botnetzes und somit können nur wenig Einsichten auf ihre Größe und Struktur gewonnen werden. Wenn das Netzwerk jedoch nur oft genug “gemessen” würde und diese Messpunkte zusammengeführt werden könnten, liesse sich dennoch die Struktur aufdecken.

Diese Erkenntnis bietet auch einen Ansatz zur Verteidigung: Da DGAs und Proxy Layers automatisierte Prozesse sind, werden i.d.R. viele Komponenten wiederverwertet, d.h. konsekutive Domain Namen nutzen den gleichen Registrar, Authoritative DNS, oder zeigen auf die gleichen Netzwerkbereiche. In dem wir das Ökosystem aller registrierten Domainnamen in der Welt beobachten, können wir diese Gemeinsamkeiten und Verhaltensauffälligkeiten algorithmisch finden, und damit frühzeitig Domainnamen, die z.B. für Malware genutzt werden (sollen), entdecken.

Eine Platform für DNS Auswertung

Unser Team bekommt tägliche Updates von über 1100 Top Level Domänen (wie z.B. .com, .net usw.) welche Domainnamen aktuell registriert sind. Durch spezielle parallelisierte Crawler können wir alle registrierten Domainnamen in nur wenigen Stunden aufrufen, und erhalten so mehrfach am Tag eine Übersichtskarte über den aktuellen Stand der “Verdrahtung” der Domainnamen im Internet. Wenngleich eine einzelne solche Karte des Internets einige hundert Gigabyte umfasst und damit eine längerfristige Beobachtung des Internets nicht skalieren würde, ist das Gros des Internets relativ statisch, und die meisten DNS Einträge bleiben auch über längere Zeit identisch. Indem diese Crawls inkrementell und Redundanz-bereinigt gespeichert werden, wird das Beobachten von Entwicklungen von Domainnamen, sowie das Suchen nach bestimmten Mustern auch über längere Zeit möglich.

Ziel dieses Projektes ist zum Einen auf Basis von Open Source Projekten wie Hadoop und Cassandra ein Datenspeicherungsformat zu entwickeln, welches diese Messwerte möglichst Ressourcen-schonend speichert und dabei gleichzeitig für Algorithmen so effizient wie möglich erschließbar hält. Hierbei soll es möglich sein, auf den Daten die Entwicklung von Gruppen von Einträgen im Laufe der Zeit zu verfolgen, sowie auf den Daten mit komplexeren (Clustering-)Algorithmen zu rechnen.

Dieses System werden wir im zweiten Schritt einsetzen, um in einer Use Case Studie die Struktur von ausgewählten Malware-Ökosystemen zu beleuchten und zu demonstrieren, dass sich ausgewählte Formen der Cyberkriminalität (Botnetze, Phishing) durch eine algorithmische Suche nach bekannten Modus Operandi der Kriminellen (welche Registrar, Systeme usw. werden eingesetzt, wie werden Systeme durchgetauscht) entdecken lassen.

Kontakt:
Prof. Christian Dörr, christian.doerr@hpi.de
Fachgebiet Cybersecurity - Enterprise Security