Clusteranalyse von Malware basierend auf Installationsgraphen

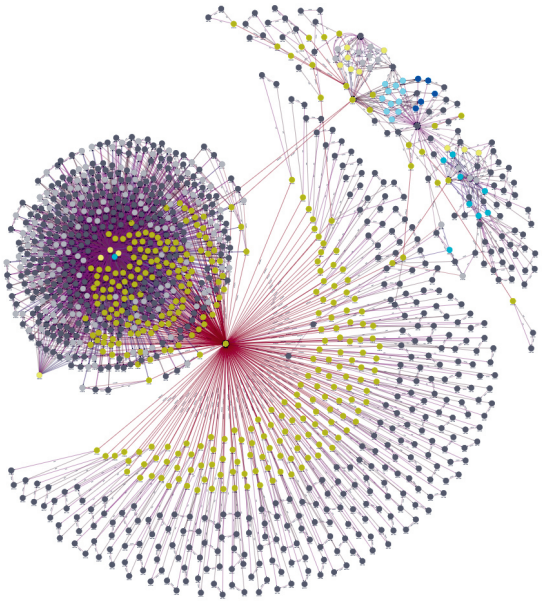
Studiengang: BSc in Informatik | Vertiefung: IT-Security
Betreuer: Dr. Endre Bangerter, Jonas Wagner
Experte: Dr. Igor Metz

Im Jahr 2015 wurden 431 Millionen neue Malware-Varianten entdeckt. Um Abwehrmechanismen entwickeln zu können, braucht es Methoden die automatisiert Malware-Samples klassifizieren. Die Praxis zeigt, dass sich Malware mit gleichem Ziel ähnlich verhält. Wir definieren Verhalten als die Entitäten in einem System und deren Interaktion. Gestützt auf dieser Beobachtung, stellen wir einen Graph-basierten Ansatz vor. Dieser erlaubt es, Malware nach Verhalten zu klassifizieren.

Ausgangslage

Da die manuelle Analyse von Malware nicht skaliert, entwickelt die IT-Security-Branche automatisierte Technologien und Prozesse.

Eine Vielzahl von Techniken und Technologien beschäftigen sich mit dem Laufzeitverhalten von Programmen. Nahezu alle Informationen die zu einer Rekonstruktion des Verhaltens benötigt werden, sind im physikalischen Speicher abgelegt. Aus diesem Grund hat des Security Engineering Lab der Berner Fachhochschule das Memory-Tracing-System KAN entwickelt. KAN überwacht den Speicher und legt bei bestimmten Aktionen eine Kopie an. Aus jeder Kopie extrahiert KAN danach Informationen über den Zustand des Systems. Eine zeitlich geordnete Sequenz dieser Zustandsinformationen nennen wir Memory-Trace. Aus Traces erstellen wir gewichtete, attributierte und gelabelte Graphen, sogenannte Installationsgraphen.



Ein Installationsgraph der Malware-Familie Cosmicduke

Um das charakteristische Laufzeitverhalten von Malware zu beschreiben, wurden bereits Graph-basierte Ansätze entwickelt. Von einem bekannten ausgehend, stellen wir in unserer Arbeit einen neuen Ansatz vor. Weiter haben wir die Hypothese geprüft, ob Installationsgraphen geeignet sind Malware nach Familie zu clustern.

Methodik

Unser Ansatz berechnet Ähnlichkeiten auf Basis von Attributen und clustert damit Installationsgraphen.

Zwei Installationsgraphen sind ähnlich, wenn sie Substrukturen teilen. Solche treten auf, wenn zum Beispiel eine Malware mit dem Ziel Administratorrechte zu erlangen, den Flashplayer verändert und dann startet. Substrukturen finden wir dank geschicktem Vergleichen von Attributen und Labels. Basierend auf diesen, gruppieren wir danach die Graphen mit einem Clusteralgorithmus. Als letzten Schritt bewerten wir die Qualität der gefundenen Cluster. Dazu vergleichen wir unsere mit bereits bekannten Clustern. Eine solche Clusterung ist die Zuordnung von Malware zu einer Familie.

Resultate

Wir haben gezeigt, dass Installationsgraphen eine geeignete Codierung von Malwareverhalten sind, die zu sinnvollen Clustern führen.

Um neues Verhalten zu verstehen, müssen in Zukunft nur noch wenige Malware-Samples genauer untersucht werden. Diese Fokussierung bedeutete, dass der Analyseprozess besser skaliert. Der entwickelte Ansatz ist erweiterbar und ermöglicht somit eine kontinuierliche Verbesserung des Analyseprozesses. Weiter haben wir bereits jetzt ähnliches Verhalten zwischen Malware-Familien, sowie Substrukturen innerhalb der Familien, gefunden. Diese ersten Erkenntnisse können nun genauer untersucht werden.

Mischa Lehmann

Sebastian Philipp Plattner