MAL2

MAchine Learning detection of MALicious content

Programm / Ausschreibung	IKT der Zukunft, IKT der Zukunft, IKT der Zukunft - 6. Ausschreibung (2017)	**Status**	abgeschlossen
Projektstart	01.01.2019	**Projektende**	31.12.2020
Zeitraum	2019 - 2020	**Projektlaufzeit**	24 Monate
Keywords	Machine Learning; Neural Networks; Malware Analysis; Fake-Shop Detection; Cybercrime Prevention		

Projektbeschreibung

Neuronale Netze sind die treibende Kraft im Bereich des Machine Learnings und zeigen sich für eine Vielzahl an Erfolgen in den unterschiedlichsten Anwendungsgebieten verantwortlich. MAL2 beabsichtigt Deep Neural Networks und Unsupervised Learning zur automatisierten Detektion von a) betrügerischen Fake-Shops und b) schädlichen Android Apps (PHAs) einzusetzen und somit zur Verbesserung der Cyberkriminalitätsprävention beizutragen.
Online-Shopping ist alltäglich geworden, bereits 61,6% der Österreicher nutzen diese Form des Einkaufens. Der Umsatz der umsatzreichsten 250 Onlineshops in Österreich lag 2016 bei 2,3 Milliarden Euro, was einem Wachstum von rund 9 Prozent gegenüber dem Vorjahr entspricht. Kunden durch betrügerische Online Shops ‚abzuzocken' ist ein rasant wachsendes Cyberkriminalitätsdelikt. Seit Juli 2013 leistet der Internet Ombudsmann (ÖIAT) Aufklärungsarbeit und trägt mit einer Blacklist auf dem Portal "Watchlist Internet" zur Prävention bei. Betrügerische Online-Shops zu entlarven ist eine zeitintensive Aufgabe und erfordert vor allem großen manuellen Aufwand, da oft Dutzende oder mehr dieser Kopien zur gleichzeitig existieren. Jede Woche werden über150 neue gefälschte Online-Shops zur manuellen Verifizierung eingereicht. MAL2 erarbeitet eine Lösung zur Klassifizierung von Fake-Shops auf Basis von strukturellen Ähnlichkeitsmerkmalen und ermöglicht somit Cybersquatting durch Machine Learning zu erkennen. Mit über 2 Milliarden aktiven Geräten pro Monat (Tablets, Smartphones, etc.) ist Android das weltweit weitverbreitetste Betriebssystem für Mobilgeräte. Alleine 2016 wurden vier Millionen neue Malware-Programme für diese Plattform veröffentlicht. Der Marktanteil von Exploits, gemessen über alle Plattformen hinweg, beträgt 21% - Android ist somit die zweihäufigste adressierte Plattform für Exploit-Angriffe. Im vierten Quartal 2016 hatten 0,71% aller Geräte Schadsoftware (PHAs) installiert. Ziel des MAL2 Projekts ist es, ein künstliches neuronales Netz zu trainieren, und die Möglichkeiten der Erkennung aufkommender Angriffsmuster und deren Nachvollziehbarkeit zu evaluieren. Ein Referenzdatensatz betrügerischer Fake-Shops ist bislang inexistent. Da die Aussagekraft neuronaler Netze stark von der Qualität der zugrundeliegenden Trainingsdaten, sowie der Wahl der geeigneten Granularität der extrahierten Merkmale abhängt, wird MAL2, zum Training der Modelle, Ground-Truth-Datensätze in beiden Anwendungsgebieten erarbeiten und veröffentlichen. Ziel des MAL2 Projektes ist es (i) ein Open-Source-Framework zu entwickeln, welches funktionelle Unterstützung in der Datenextraktion, Featureerstellung, dem Training der NNs sowie Analyse der Ergebnisse ermöglicht (ii) eine performante Ausführung der Komponenten auf Hadoop und GPU-Clustern ermöglicht und (iii) sowie die Referenzdatensätze, die extrahierten Features sowie die trainierten NNs in beiden

Anwendungsdomänen zu veröffentlichen. Die Projektergebnisse werden auf Watchlist-Internet, in Form von Demonstratoren die eine Live-Detektion und Risikoabschätzung von Apps und Online-Shops ermöglichen, veröffentlicht und trägt unter anderem zu einer gesteigerten Bewusstseinsbildung für Maßnahmen zur Cyberkriminalitätsprävention und Sensibilisierung bei.

Abstract

Neural Networks are a dominant force in machine learning and are responsible for the massive momentum in deep learning in numerous application domains. MAL2 will apply Deep Neural Networks and Unsupervised Learning to advance cybercrime prevention by
a) automating the discovery of fraudulent eCommerce and b) evaluating the capabilities of detecting Potentially Harmful Apps (PHAs) in Android operating systems.
Online shopping is commonplace, with 61.6% of Austrians already using this form of commerce. The turnover of the top 250 online shops in Austria in 2016 was € 2.3 billion, which corresponds to growth of around 9 percent compared to the previous year. Ripping of customers through fraudulent eCommerce shops is a rapidly growing area in cybercrime. Since July 2013, the Internet Ombudsman (ÖIAT) offers preventive information and maintains a blacklist on the "Watchlist Internet" portal. Exposing such fake offerings however is a labour intensive, manual task as often, dozens or more of these copies exist at the same time - every week more than 150 new fake online-shops are entered for manual verification. MAL2 provides means for advancing the automation and detection of fake-shop cybersquatting through machine learning technologies by classifying sites based on their structural similarity.
With over two billion monthly active devices, the Android operating system for tablets, phones and smart devices it by far the most widespread mobile operating system in the world. Four million new malware programs were released for this platform in the year 2016. The total market share of exploits that target the Android platform is 21% which makes it the second most targeted platform for running exploit attacks. By Q42016 0.71% of all devices had potentially harmful applications (PHAs) installed. The goal of the project is to train a Neural Network to evaluate the discoverability and explainability of upcoming attack patterns.
Classification capabilities of Neural Networks are heavily reliant on the quality of the underlying datasets, and subsequently dependent even more on the granularity of extracted features. Up to date no web-archive dataset of fraudulent eCommerce sites has been collected and released. MAL2 will collect/harvest and curate two large-scale Ground-Truth dataset existing of a) malware/benign applications and b) web-archives of fake-shops, to train its machine learning detection models in the application domains.
Currently there is a lack of technology supporting an integrated solution of large-scale feature extraction and Neural Network training. The goal of the MAL2 project is (i) to release Open Source framework which provides integrated functionality along the required pipeline – from data extraction, feature composition up to Neural Network training and analysis of results (ii) to execute its components at large-scale within Hadoop and GPU cluster support and (iii) to publish the harvested Ground-Truth dataset, the extracted features as well as the trained Neural Network in both application domains on open data platforms. To visualize the projects results and to raise awareness for cybercrime prevention in the general public, two demonstrators are deployed at Watchlist Internet that allow live-inspection on the trustworthiness of eCommerce sites and Android Apps.

Projektkoordinator

- AIT Austrian Institute of Technology GmbH

Projektpartner

- X-Net Services GmbH
- IKARUS Security GmbH
- Österreichisches Institut für angewandte Telekommunikation
- Kompetenzzentrum Sicheres Österreich (KSÖ)