idw - Informationsdienst Wissenschaft
Nachrichten, Termine, Experten

Press release

Technische Universität Berlin
Stefanie Terp

11/02/2022
http://idw-online.de/en/news803986

Research results
Economics / business administration, Electrical engineering, Information technology, Law
transregional, national

BIFOLD: Cybersicherheit auf dem Prüfstand

Maschinelles Lernen in der Sicherheitsforschung birgt subtile Fallstricke

Cybersicherheit ist ein zentrales Thema der digitalen Gesellschaft und spielt sowohl im kommerziellen wie auch privaten Kontext eine wesentliche Rolle. Maschinelles Lernen (ML) hat sich in den letzten Jahren als eines der wichtigsten Werkzeuge zur Analyse sicherheitsrelevanter Probleme herauskristallisiert. Eine Gruppe europäischer Forscher*innen der TU Berlin, der TU Braunschweig, des University College London, des King's College London, der Royal Holloway University of London und des Karlsruher Instituts für Technologie (KIT)/KASTEL Security Research Labs unter der Leitung von BIFOLD-Forschern der TU Berlin konnte jedoch zeigen, dass diese Art der Forschung oft fehleranfällig ist. Ihre Veröffentlichung: „Dos and Don'ts of Machine Learning in Computer Security“ über Fallstricke bei der Anwendung von Maschinellem Lernen in der Sicherheitsforschung wurde auf dem renommierten USENIX Security Symposium 2022 mit einem Distinguished Paper Award ausgezeichnet.

Maschinelles Lernens (ML) hat in einer Vielzahl von Anwendungsbereichen, wie zum Beispiel der Bilderkennung und der Verarbeitung natürlicher Sprache, zu großen Durchbrüchen geführt. Dieser Erfolg wirkt sich auch auf die Cybersicherheit aus: Nicht nur kommerzielle Anbieter werben damit, dass ihre von künstlicher Intelligenz (KI) gesteuerten Produkte effizienter und effektiver als bisherige Lösungen sind. Auch viele Forscher*innen setzen diese Technik ein, da Algorithmen den traditionellen Methoden oft weit überlegen zu sein scheinen. So wird maschinelles Lernen zum Beispiel auch eingesetzt, um neue digitale Angriffstaktiken zu erlernen und die Abwehrmaßnahmen an diese Bedrohungen anzupassen.
„In dem Paper liefern wir eine kritische Analyse des Einsatzes von ML in der Cybersicherheitsforschung", beschreibt Erstautor Dr. Daniel Arp, Postdoc an der TU Berlin: „Zunächst identifizieren wir häufige Fallstricke bei der Konzeption, Implementierung und Evaluierung von lernbasierten Sicherheitssystemen.“ Ein Beispiel für solche Probleme ist die Verwendung nicht repräsentativer Daten. Also Datensätze, bei denen die Anzahl der Angriffe im Vergleich zu ihrer Häufigkeit in der Realität überrepräsentiert ist. ML-Modelle, die auf solchen Daten trainiert wurden, können sich in der Praxis als unbrauchbar erweisen. Im schlimmsten Fall könnte sich sogar herausstellen, dass sie außerhalb einer experimentellen Umgebung gar nicht funktionieren oder zu Fehlinterpretationen führen.
In einem zweiten Schritt führten die Forscher eine Prävalenzanalyse auf der Grundlage der identifizierten Probleme durch, bei der sie 30 Beiträge von hochrangigen Sicherheitskonferenzen untersuchten, die zwischen 2010 und 2020 veröffentlicht wurden. „Zu unserer Besorgnis mussten wir feststellen, dass diese Fallstricke selbst in sorgfältig durchgeführter Spitzenforschung weit verbreitet sind“, sagt BIFOLD Fellow Prof. Dr. Konrad Rieck von der TU Braunschweig.

Wo moderne Cybersecurity-Ansätze ins Straucheln kommen

Auch wenn diese Ergebnisse bereits ein alarmierendes Signal waren - die möglichen Folgen waren zunächst unklar. In einem dritten Schritt haben die Forscher*innen daher anhand von vier konkreten Fallstudien mit Beispielen aus der

Literatur gezeigt, wie und wo diese identifizierten Probleme zu unrealistischen Ergebnissen und Interpretationen von ML-Systemen führen.

Eine der untersuchten Fallstudien beschäftigte sich mit der Erkennung mobiler Schadsoftware, sogenannter Malware. Aufgrund der großen Anzahl neuer gefährlicher Software für mobile Geräte, haben herkömmliche Antiviren-Scanner oft Probleme, mit der Schadsoftware Schritt zu halten und bieten nur eine schlechte Erkennungsleistung. Um dieses Problem in den Griff zu bekommen, haben Forscher*innen lernbasierte Methoden vorgeschlagen und entwickelt, die sich automatisch an neue Malware-Varianten anpassen können.
„Leider wurde die Leistung der lernbasierten Systeme in vielen Fällen überschätzt. Da es keine öffentlich zugänglichen Lern-Datensätze von Unternehmen gibt, nutzen Forscher*innen meist eigene Datensätze und führen dazu verschiedene Quellen zusammen", erklärt Dr. Daniel Arp. „Diese Zusammenführung der Lern-Datensätze aus verschiedenen Quellen führt jedoch zu einer Verzerrung der Stichprobe: Apps aus den offiziellen App Stores der Smartphonehersteller bergen tendenziell weniger Sicherheitsrisiken als Apps, die aus alternativen Quellen mit geringeren Sicherheitsstandards stammen. Im Ergebnis konnten wir zeigen, dass moderne Cybersecurity-Ansätze dazu neigen, sich bei der Erkennung von Schadsoftware auf Merkmale zu konzentrieren, die auf die Quelle der App zurückzuführen sind, anstatt reale Malware-Merkmale zu identifizieren. Dies ist nur eines von vielen Beispielen des Papers, die zeigen, wie ein kleiner Fehler bei der Zusammenstellung der Lern-Datensätze, schwerwiegende Verzerrungen im Ergebnis herbeiführt und das gesamte Experiment beeinflussen kann.“

Die Probleme bei der Anwendung von ML-Methoden in der Cybersicherheit werden durch die Notwendigkeit, in einem feindlichen Kontext zu arbeiten, noch verschärft. Mit ihrer Veröffentlichung hoffen die Forscher*innen, das Bewusstsein für potenzielle Fehlerquellen im experimentellen Design zu schärfen und diese wenn möglich zu verhindern.

Publikation:

Daniel Arp, Erwin Quiring, Feargus Pendlebury, Alexander Warnecke, Fabio Pierazzi, Christian Wressnegger, Lorenzo Cavallaro, Konrad Rieck: Dos and Don'ts of Machine Learning in Computer Security,
https://www.usenix.org/system/files/sec22-arp.pdf

Weitere Informationen erteilt Ihnen gern:
Dr. Daniel Arp
Tel.: 0049 (0)30 314-78621
d.arp@tu-berlin.de