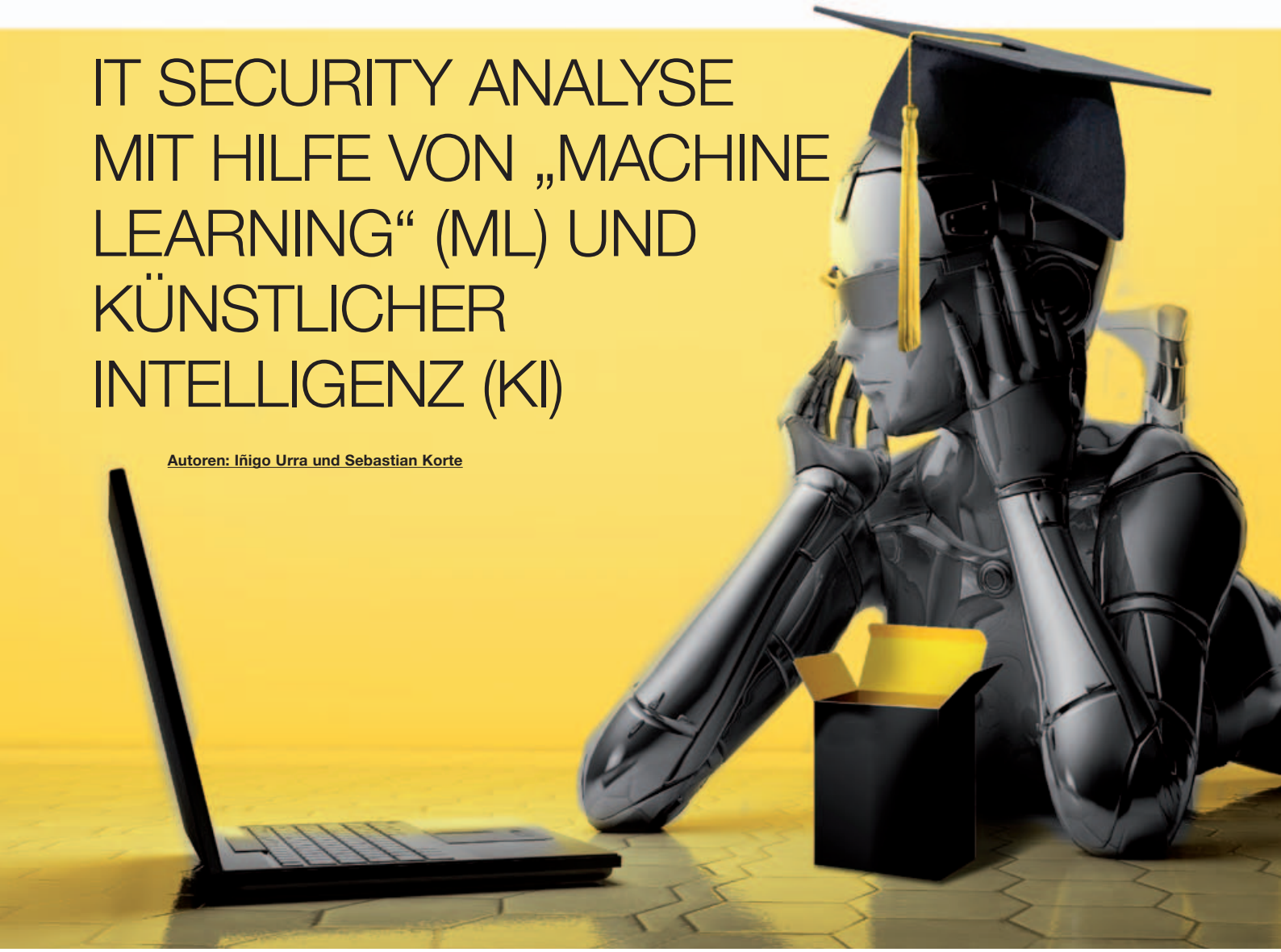
IT SECURITY ANALYSE MIT HILFE VON „MACHINE LEARNING“ (ML) UND KÜNSTLICHER INTELLIGENZ (KI)

Autoren: Iñigo Urra und Sebastian Korte



Maschinelles Lernen und Informationssicherheit

In den letzten Jahren sind der elektronische Informationsaustausch und damit die Erzeugung von Daten exponentiell gestiegen. Es sind mehr Daten generiert worden als je zuvor. Einer der Gründe für diese Revolution ist das „Internet der Dinge“, das dazu beigetragen hat, tausende von verschiedenen Geräten miteinander zu verbinden.

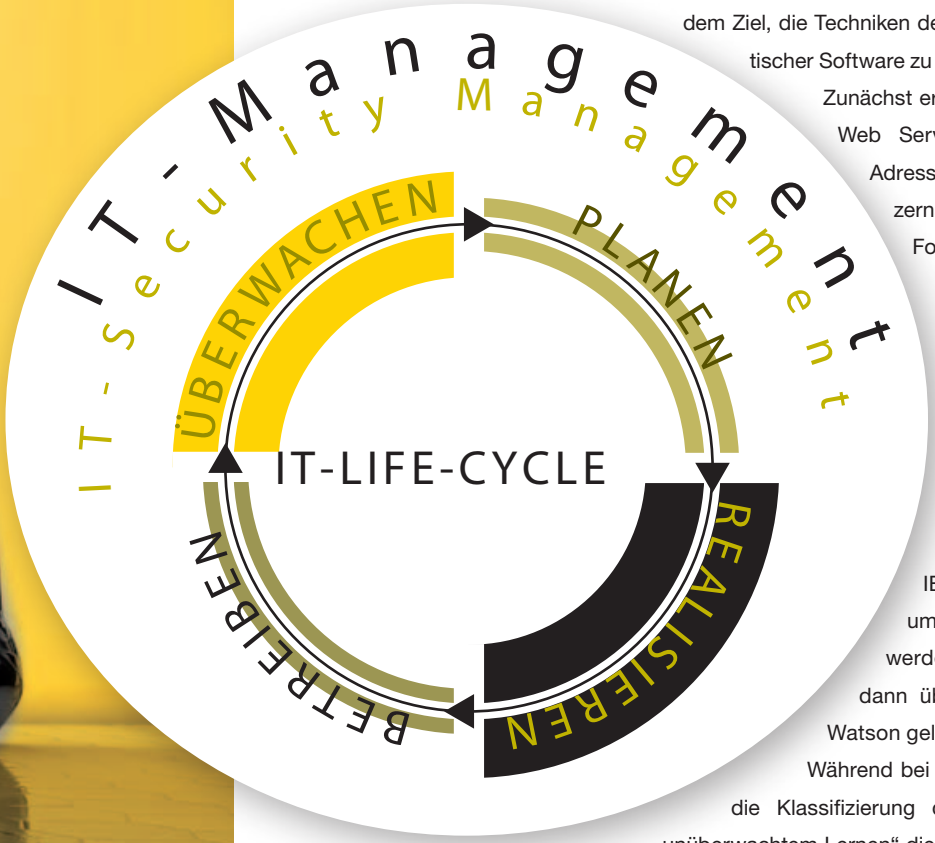
Unternehmen nutzen neue Technologien in ihrer täglichen Arbeit, die auch in kleinen Firmen längst selbstverständlich sind. In großen Firmen wird diese Entwicklung noch offensichtlicher, wo „Business Intelligence“ hilft, bessere Entscheidungen zu treffen, Menschen mit modernen Kommunikationsmitteln länderübergreifend zusammenarbeiten und auch vertrauliche Informationen teilen.

Die Menge der Daten stellt die Unternehmen vor neue Herausforderungen. Um die massenhaft generierten Informationen auch sinnvoll nutzen zu können, benötigen wir Methoden und Werkzeuge, um diese auswerten und analysieren zu können. Eine manuelle Analyse ist längst nicht mehr möglich und die verfügbaren Werkzeuge setzen meist erhebliche personelle und wirtschaftliche Ressourcen voraus. Das gilt auch für den Bereich der Informationssicherheit. Insbesondere für die Analyse von sicherheitsrelevanten Daten, die mit Hilfe von SIEM (Security Information and Event Management) Werkzeugen durchgeführt wird, um Sicherheitsvorfälle erkenn zu können, gilt, dass gut ausgebildete Analysten benötigt werden, um diese Daten bewerten zu können.

ConSecur untersucht in einem F+E Projekt, wie ML (Machine Learning) und KI (Künstliche Intelligenz) uns dabei unterstützen können, immer mehr Daten immer schneller und effizienter zu analysieren, um Sicherheitsvorfälle zuverlässiger identifizieren und bekämpfen zu können.

Das Projekt betrachtet zwei methodische Ansätze, um das Ziel zu erreichen. Im ersten Schritt werden Aufgaben automatisiert, etwa die Erstellung von Angriffserkennungsregeln für das SIEM-System, um die Security Analysten von Routinetätigkeiten zu entlasten und nur noch die wirklich wichtigen Arbeiten erledigen zu lassen. Der zweite Schritt ist die Reduzierung von falschen Alarmmeldungen (sog. false positives), so dass sich der Analyst nur noch um relevante Bedrohungen kümmern muss.

Bild: fs Quelle: Kovalenko-fotolia



Für das Erkennen von unbekannten Angriffen wurde die Methode des sog. „Nicht überwachten Lernens“ betrachtet. Dabei werden die Daten, die von einem SIEM-System verarbeitet werden (Log-Daten) anhand ihrer Attribute einer Klasse (z.B. „normal“ und „verdächtig“) zugeordnet. Mittels verschiedener Lernmethoden (Clustering, Anomalie-Erkennung etc.) erlernt das System dann nach und nach, welche Log-Einträge auf gewünschte Kommunikation hinweisen und welche einen Angriff als Hintergrund haben. Diese Erkenntnis kann dann als Regel an das SIEM-System weitergegeben werden, welches diese Angriffe dann erkennen kann. Um bekannte Angriffsmuster zu erkennen und false positives auszuschließen wurde die Methode des „überwachten Lernens“ angewendet. Mit Hilfe von ausgewählten Trainingsdatensätzen überwacht das System die eingehenden Logeinträge und vergleicht diese mit den vorgenommenen Klassifizierungen. Auf diese Weise können notwendige Änderungen der Klassifizierungen erkannt und per Regeländerungen an das SIEM-Tool weitergegeben werden. Damit kann die Regelbasis eines SIEM-Systems automatisiert optimiert und schneller an neue Bedrohungen angepasst werden.

Die Umsetzung mit Watson

Im weiteren Verlauf soll der zuvor beschriebene Ansatz durch das von IBM entwickelte Cognitive System Watson umgesetzt werden. Watson ist ein System aus dem Bereich der Künstlichen Intelligenz, dass mit dem Ziel, die Techniken der künstlichen Intelligenz mit analytischer Software zu verbinden, erschaffen wurde.

Zunächst erzeugen wir mit Hilfe eines Apache Web Servers ausreichend Logdaten (IP-Adressen, angeforderte Ressourcen, Nutzernamen etc.), und geben diese im Folgenden an Watson weitergeben.

Nachdem die Daten in Watson Analytics geladen wurden, soll Watson mittels „überwachtem Lernen“ – innerhalb der Logdaten Muster lernen, um später Anomalien erkennen zu können. Dies geschieht über eine Anwendung, die in der für Watson von IBM bereitgestellten Entwicklungsumgebung „Blue Mix“ implementiert werden muss. Diese Anwendung kann dann über eine Schnittstelle, auf die in Watson geladenen Daten zugreifen.

Während bei der oben beschriebenen Methode die Klassifizierung durch uns erfolgt, kann mittels „unüberwachtem Lernen“ die Klassifikation durch Watson erfolgen. Hierbei werden Wahllos Daten ausgewählt und auf Gemeinsamkeiten mit bereits vorhanden Mustern untersucht, bis alle Daten gemäß ihrer Eigenschaften zerlegt und in Gruppen (Clustern) aufgeteilt wurden. Die Gruppen entsprechen dann den jeweiligen Klassifikationen. Neue Daten können dann mit denen aus den verschiedenen Gruppen/Klassifikationen verglichen werden.

Die bisherigen Versuche haben gezeigt, dass sich die Lernphasen durchaus aufwendig gestalten. Jedoch sind diese die wesentlichen Erfolgsfaktoren für ein späteres Funktionieren des Prototypens. In einem nächsten Schritt wird Watson dann lernen müssen, wie es mit den erkannten Anomalien umgehen soll.

Iñigo Urra ist SIEM Berater bei ConSecur

Sebastian Korte ist Entwickler bei ConSecur