21.05. 2019

Deutsch-französisches IT-Sicherheitslagebild

2.Edition

Bundesamt
für Sicherheit in der
Informationstechnik

Vorwort von Arne Schönbohm – Präsident des Bundesamts für Sicherheit in der Informationstechnik

Die Digitalisierung stellt Staaten vor Herausforderungen, die nicht allein auf der nationalen Ebene gelöst werden können. Sie steht für den Fortschritt in der Gesellschaft und der Weltwirtschaft. Auf der anderen Seite bringt die Digitalisierung auch Herausforderungen im Feld der Cyber-Sicherheit mit sich, die per se internationale Lösungen erfordern. Daher wird die starke Kooperation der nationalen Cyber-Sicherheitsbehörden immer unverzichtbarer.

Vor diesem Hintergrund basiert die Beziehung von ANSSI und BSI auf gegenseitigem Vertrauen, Respekt und einer langfristigen Partnerschaft, die alle technischen Herausforderungen einer sicheren und erfolgreichen Digitalisierung abdeckt. Die Basis unserer täglichen Zusammenarbeit ist der regelmäßige Austausch unserer Experten zu Themen der Cyber-Sicherheit, Standardisierung, Zertifizierung und Kryptografie.

Im Juli 2018 veröffentlichten ANSSI und BSI ihr erstes gemeinsames IT-Sicherheitslagebild, das die Bedrohung durch Ransomware und Cryptocurrency-Crime und Maßnahmen gegen diese beleuchtete. Die zweite Edition bietet nun ein Update zum Thema Cryptocurrency-Crime und beleuchtet zudem das wachsende Feld der Künstlichen Intelligenz, dessen Überschneidungen mit Cyber-Sicherheit zumeist unterschätzt werden. Die Analyse dieser sich schnell entwickelnden Thematik zeigt die Notwendigkeit der engen Synchronisation unser gewonnenen Erkenntnisse, um sowohl aktuelle als auch zukünftige Herausforderungen erfolgreich zu meistern.

Vorwort von Guillaume Poupard – Generaldirektor der Agence nationale de la sécurité des systèmes d'information

Unsere Behörde feiert in diesem Jahr ihren 10. Geburtstag. Über die letzten 10 Jahre ist das BSI unser ältester und nahe stehendster Partner unter anderen in der Zusammenarbeit geblieben. Diese zweite Edition des Deutsch-französischen IT-Sicherheitslagebilds ist sowohl eine Bestandsaufnahme der vergangenen als auch ein lebendiger Beweis der engen Zusammenarbeit von ANSSI und BSI. Diese Kooperation wird auf einer höheren Ebene durch die Unterzeichnung des bilateralen Vertrags von Deutschland und Frankreich vom 22.01.2019 gespiegelt, welcher auf einer Ebene mit dem vor 55 Jahren unterzeichneten Élysée-Vertrag steht.

Diese Veröffentlichung hätte nicht erarbeitet werden können ohne das starke Fundament der Zusammenarbeit, die durch den täglichen Austausch der beiden Behörden belegt wird, sei es zur Zertifizierung, auf der technischen, operationellen Ebene oder zu Forschungsergebnissen. Über den Austausch hinaus hat diese Zusammenarbeit dazu beigetragen, dass beide Partner ihre Erwartungen und Ziele übertroffen haben.

Die in diesem Lagebild adressierten Themen sind Beispiele für die bevorstehenden Bedrohungen und Möglichkeiten. Verbrechen in Verbindung mit Kryptowährungen sind ein Resultat der tiefgreifenden Veränderungen in der Gesellschaft, die durch die Digitalisierung hervorgebracht werden. Diese führt zu einem Paradigmenwechsel der Angriffsmöglichkeiten, die in erheblichem Umfang mit kriminellen Absichten durchgeführt werden. Künstliche Intelligenz wird einen ebenso großen Wandel herbeiführen und gemeinsam mit unseren Kollegen vom BSI werden wir die Möglichkeiten nutzen Sicherheit in der Informationstechnik zu gestalten.

Inhalt

Einleitung

In der ersten Edition des Deutsch-französischen IT-Sicherheitslagebilds, legten ANSSI und BSI den Fokus auf Angreiferaktivitäten in Zusammenhang mit Ransomware und Cryptocurrency-Mining, was eine wachsende Bedrohung für Privatpersonen, Betreiber kritischer Infrastrukturen und kleine und mittelständische Unternehmen darstellte.

In dieser zweiten Edition wird die Entwicklung im Feld der Kryptowährungen aufgegriffen, indem auch Angriffsmethoden auf die zu Grunde liegende Blockchain-Technologie erläutert werden. Weiterhin bietet das Lagebild 2019 eine Einführung zu Kernprinzipien der Künstlichen Intelligenz (KI) und insbesondere ihren Implikationen für die Zukunft der IT-Sicherheit.

Cryptocurrency-Crime

Cryptocurrency-Crime war ein zunehmender Trend in 2017. Der enorme Kursanstieg bereitete die Basis für zahlreiche technische und nicht-technische Angriffe auf die zu Grunde liegende Blockchain-Technologie und den Handel mit Kryptowährungen. In dem folgenden Update zu der Ausarbeitung von 2018 zur Lage im Feld Kryptowährungen werden einige Beispiele für die kriminelle Ausnutzung der bestehenden Mechanismen benannt.

Die Volatilität der neuen Pseudo-Währungen wurde in 2018 allgegenwärtig, was in den Schwankungen der Gesamtmarktkapitalisierung deutlich wurde. Anfang 2019 fiel die Kapitalisierung der größten Kryptowährung auf einen Bruchteil des der Zahlen vom November 2017. Daher ließ auch das kriminelle Interesse nach. Dennoch werden durch die gegebenen Beispiele grundsätzliche Sicherheitsaspekte der Blockchain-Technologie angesprochen und es wird deutlich wie eine transparente und sichere Konzeption Opfer ihrer eigenen Mechanismen werden kann.

Künstliche Intelligenz – Risiko und Hoffung für die IT-Sicherheit

Künstliche Intelligenz ist nicht nur ein interessanter und sich schnell entwickelnder Teil der Informationstechnologie, sondern hat bereits heute auf einen großen Teil der bestehenden Anwendungen enorme Auswirkungen. Diese spielen eine wichtige Rolle für die wirtschaftlichen Entwicklungen.

In der wissenschaftlichen Literatur sowie in alltäglichen Nachrichten wird eine wachsende Zahl von Erfolgsgeschichten veröffentlicht, wie KI im täglichen Leben Einzug hält. Es gibt eine Vielzahl von Beispielen für die Anwendung von KI. Internetnutzer erhalten Suchergebnisse, die durch Maschinelles Lernen (ML) erzeugt wurden, Versicherungen bestimmen ihre Tarife, Banken entscheiden über Aktienkäufe und -verkäufe. Sogar die Polizei nutzt die “vorhersagende Analyse” (Predictive Analytics) um zu bewerten, in welchen Regionen Einbrüche wahrscheinlich sind.

Schwachstellen der Künstlichen Intelligenz könnten daher zu einer großen Angriffsfläche führen. Um das Ausmaß des zu erwartenden Einflusses zu beleuchten, werden hier einige grundsätzliche Aspekte der Verwundbarkeit künstlicher Intelligenz anhand von Beispielen präsentiert. Weiterhin wird KI als Waffe für Cyberattacken sowie ihre Anwendbarkeit zu Verteidigungszwecken beleuchtet.

Das gemeinsame Lagebild kommt zu dem Schluss, dass KI bereits heute Einfluss auf die IT-Sicherheit hat. Sie wird vielseitig im privaten und wirtschaftlichen Umfeld eingesetzt und in Zukunft eine bedeutende Rolle für die Informationssicherheit spielen. Daher ist eine hohe Aufmerksamkeit in Bezug auf die weitere Entwicklung ratsam. Eine hohe Expertise bzgl. der Wirkungsweisen und die Beobachtung von Hardware- und Software-Entwicklungen sowie von Lieferketten ist daher von essentieller Bedeutung.

Die Lage im Cryptocurrency-Crime: Update

Im ersten Deutsch-französischen IT-Sicherheitslagebild stand Cryptocurrency-Crime im Fokus. Der enorme Kursanstieg vieler Kryptowährungen Ende 2017 und Anfang 2018 befeuerte die Entwicklung in diesem neuen Feld des Cybercime, bspw. durch Cryptojacking.

In der Zeit von Januar 2017 bis Januar 2018 stieg der Kurs des Bitcoin von 900 auf 13.000 EUR (+1440%). Dieser Trend hielt jedoch nicht an, die Spekulationsblase platzte und die Kurse von Kryptowährungen fielen massiv im ersten Quartal 2018. Der Kurs des Bitcoin bspw. sank von rund 16.000 EUR (Januar 2018) auf 3.500 EUR (März 2019), was einen Verlust von 80% des Kurspreises bedeutet. Mögliche Gründe für den Kursfall von Kryptowährungen und die hohe Volatilität in 2018 sind die übersteigerte mediale Aufmerksamkeit – hierdurch das Anziehen unerfahrener Investoren -, internationale Bestrebungen für Rechtsgebung, wie Steuer- und Geldwäschegesetze, die Diversifizierung des Kryptowährungsmarkts, Unsicherheiten auf Grund von Blockchain-Forks, die Kritik am hohen Energieverbrauch des Minings und Berichte über Sicherheitsvorfälle. Diese Faktoren könnten Investoren beeinflusst haben, die ihr Kapital zurückzogen und einen Dominoeffekt hervorruften.

Einige Monate nach dem Fall der Kurse verzeichneten ANSSI und BSI einen signifikanten Rückgang im Bereich Cryptocurrency-Crime, insbesondere beim Cryptojacking. Wie bereits im ersten Lagebild festgestellt wurde, ist Monero die am stärksten genutzte Kryptowährung beim Cryptojacking. ANSSI und BSI stellen fest, dass das Cryptojacking unprofitabel wurde und viele Schadprogramme zu diesem Zweck ihre Aktivität wegen der gesunkenen Kurse eingestellt haben [1,2].

Abb. 1: Entwicklung des Bitcoin-Euro-Kurses 01.10.2017-04.04.2019

Auch wenn einige Cryptocurrency-Crime-Trends wie Cryptojacking in 2018 bis Anfang 2019 zurückgegangen sind, haben andere Techniken an Bedeutung gewonnen. Einige werden im folgenden Abschnitt vorgestellt. [3]

Angriffsarten

Die Integrität der Blockchain wird auf theoretischer Basis durch kryptografische Protokolle sichergestellt. Dennoch gibt es Techniken, um den Mechanismus zur Validierung von Transaktionen zu umgehen, so wie die Verwendung von Drittsoftware und -diensten (Mining Pool, Kryptobörsen, "Mixer", Wallet Software, Schlüsselspeicherungssoftware), und Kryptowährungen zahlreichen Angriffen auszusetzen.

▶ 51%-Angriffe – Grenzen der Blockchain-Integrität

Der Konsensmechanismus "Proof of Work" ist fundamental, um der Blockchain Blöcke hinzuzufügen. Wenn jedoch ein Miner mehr als 50% der Gesamtrechenleistung (total hash rate) stellt, kann dieser neue Blöcke mit für ihn vorteilhaften Inhalten erstellen und als neuen Block der Blockchain hinzufügen.

1 Technik, die auf legitime und lang besuchten Webseiten eingesetzt wird. Häufig wird JavaScript zur Einbettung in die Webseite benutzt. Die Malware erzeugt im Hintergrund Kryptowährungen, während der Nutzer die Seite aufgerufen hat.

2 Marius Musch, Web-based Cryptojacking in the Wild. Chaos Communication Congress, Leipzig 2018.

3 SIX Financial Information via https://www.finanzen.net

Aufgrund der Kontrolle über die meiste Rechenleistung wird die durch den Angreifer erzeugte Kette länger als die anderer Knoten, sodass diese als valide angenommen wird. Ein 51%-Angriff eröffnet verschiedene Möglichkeiten für den Angreifer, wie das "Double Spending". Durch die Kontrolle der Rechenleistung werden angenommene Transaktionen revidiert und zu den Nicht-bestätigten zurück geführt, wodurch diese Transaktion ein zweites Mal getätigt werden kann.

***Beispiel**. 16.-19.05.2018: Bitcoin Gold, der sich im Oktober 2017 vom Bitcoin abgespalten hat, wird von einem 51%-Angriff getroffen. Der Angreifer betreibt "Double Spending" bei Kryptobörsen mit hohem Marktvolumen. Der potentielle Schaden beläuft sich auf 18 Mio. USD. In der Folge wurde der Hash-Algorithmus von Bitcoin Gold verändert.*

▶ Selfish-Mining – Bedrohung "kleinerer" Kryptowährungen

Ein Selfish-Mining-Angriff erfordert eine signifikante Gesamtrechenleistung, aber nicht notwendigerweise 51%. Er konzentriert sich auf die Belohnung des Minings, das das Proof-of-Work-Protokoll vorsieht. Dabei veröffentlicht der Angreifer Blöcke mit Verspätung, sodass er Weitere basierend auf seiner Lösung erzeugen kann. Währenddessen verschwenden andere Miner ihre Rechenleistung auf die Lösung des ersten Blocks. Hierdurch erlangt der Angreifer einen asymmetrischen Vorteil, indem er den für ihn attraktiveren Ast der Kette weiter fortsetzt. In Bezug auf Eyal and Sirer[4] ist eine Rechenleistung von 25% erforderlich für einen erfolgreichen Selfish-Mining-Angriff. Bei "kleineren" Kryptowährungen mit einer geringen Gesamt-Hashrate können 25% oder gar 51% leicht durch Cloud-Mining erreicht werden.

***Beispiel**. 13.-15.05.2018: Ein besonderer Angriff betraf Besitzer der japanischen Kryptowährung Monacoin. Der Selfish-Mining-Angriff verursachte einen Schaden von 90.000 USD, indem der Angreifer Coins zu anderen Kryptobörsen sendete, diese als Guthaben erhielt und anschließend die Transaktionen revidierte.*

▶ Cryptocurrency-Betrug – einer der wichtigsten Trends im Cryptocurrency-Crime

Neben dem Angriff auf das Konsensprotokoll werden beim Betrug Drittsoftware oder -dienste verwendet, bspw. IOTA war derartig betroffen. Über mehrere Monaten hinweg wurde ein kostenloser Dienst für die Erzeugung privater Schlüssel auf der inoffiziellen Webseite iotaseed.io angeboten. Während dieser Zeit wurden die erzeugten Zugangsdaten von den Anbietern gespeichert und schließlich genutzt, um alle IOTA-Wallets der ehemaligen Kunden zu leeren [5]. Dies ist ein Betrugsdelikte, von dem sowohl Privatpersonen als auch Geschäftsleute betroffen sein können.

▶ Angriffe auf Kryptobörsen– der profitabelste Trend

Handelsplattformen sind Dienste von Drittanbietern, die es Nutzern erlauben Kryptowährungen und traditionelle Fiatwährungen zu handeln. Diese Plattformen sind vergleichbar zum Online-Banking und stellen momentan das Hauptangriffsziel für erfahrene Angreifergruppen und individuell Agierende dar. In 2017 und 2018 wurde nicht weniger als fünf Plattformen in Teilen kompromittiert und Einlagen von dort erbeutet. Den größten Verlust verzeichnet die japanische Handelsplattform Coincheck. Die Angreifer verursachten einen Schaden von 470 Mio. EUR in der Kryptowährung NEM.

Mehreren Forschungsgruppen[6, 7] zufolge wurde der Angriff auf Handelsbörsen ein Trend in 2018 mit einem Gesamtschaden in Höhe von etwa 1 Bio. USD in Kryptowährung.

---

4 I. Eyal and E. Sirer: Majority Is Not Enough: Bitcoin Mining Is Vulnerable. Lecture Notes in Computer Science 8437, S.436–454, 2014, https://arxiv.org/pdf/1311.0243.pdf

5 McAffee: Blockchain-Threats-Report. https://www.mcafee.com/enterprise/en-us/assets/reports/rp-blockchain-security-risks.pdf. 2018.

6 Ciphertrace, Cryptocurrency Anti-Money Laudering Reports, 2018 Q3.

7 ChainAnalysis, Crypto Crime Report, Decoding increasingly sophisticated hacks, darkent markets, and scams. January 2019.

Zukünftige Bedrohungsentwicklung

Um ein Gespür für neue Cyber-Crime-Trends und -Bedrohungen zu entwickeln, ist es sinnvoll, einen Blick auf Kryptowährungstrends in der nahen Zukunft zu werfen.

In der ersten Edition des Deutsch-französischen Lagebilds erläuterten ANSSI und BSI Cryptojacking als einen Aspekt des Cryptocurrency-Crime, bei dem Rechenleistung – und hiermit auch Stromkosten – von fremden Rechnern genutzt wird, um neue Kryptowährungen zu erzeugen. Monero wird häufig bei diesem Angriffstyp verwendet, da dessen Mining-Protokoll eine hohe Anonymität garantiert. Die "Privacy Coins" (Monero, ZedCash, Verge und andere "kleinere" Kryptowährungen) werden zunehmend für kriminelle Aktivitäten genutzt, wie z.B. bei illegalen Geschäften auf Darknet-Seiten. Tatsächlich boten Plattformen wie Alphabay und Hansa (beide geschlossen in 2018) Zahlungen mit Bitcoin und Monero an. Bis 2018 waren illegale Aktivitäten bei großen Kryptowährungen pseudonym, was bedeutet, dass sie öffentlich, aber nicht nachverfolgbar waren. Mittlerweile verlassen sich Sicherheitsdienste zunehmend auf Analysetools, sodass Cyber-Kriminelle zunehmend "Privacy Coins" wie Monero auf Grund der Verschlüsselung und Nicht-Nachverfolgbarkeit bevorzugen.

Auf Grund von Schwierigkeiten bei der Mining-Konzentration (51%-Angriffe, Selfish-Mining) und auf Grund der Kritik an dem hohen Energieverbrauch entwickelten sich neue Kryptowährungen mit unterschiedlichen Mining-Protokollen. Proof-of-Stake (PoS) könnte den Proof-of-Work in den kommenden Jahren ersetzen. Neue PoS-Kryptowährungen (z.B. Tezos, EOS) bauen auf ein Mining-System, bei dem der Nutzer den Besitz einer gewissen Menge von Coins nachweisen muss, um valide zusätzliche Blöcke erzeugen zu können.

Wenn entdeckt wird, dass ein Nutzer einen falschen Block validiert, verliert er seine Anteile an der Kryptowährung. Zukünftig muss ausgewertet werden, ob Angreifer neue Techniken basierend auf dem PoS-Protokoll zum Angriff auf Drittsoftware oder -dienste entwickelt haben.

Nach mehreren Angriffen auf Handelsbörsen in 2018 wurde die gleichzeitige Nutzung mehrerer Handelsplattformen Standard, um das Risiko zu verteilen. Anstatt alle privaten Schlüssel allein in einem einfach zugänglichen Speicherplatz aufzubewahren, ist es empfehlenswert ein Protokoll für die Verbindung zu verwenden, dass es dem Nutzer erlaubt, den Schlüssel lokal zu speichern. Sollte die Dezentralisierung bei der Nutzung von Kryptobörsen weiter zunehemen, ist es wahrscheinlich, dass Cyber-Kriminelle ihre Aufmerksamkeit anderen Angriffstechniken zuwenden müssen.

Künstliche Intelligenz in der IT-Sicherheit

Die Allgegenwärtigkeit von Künstlicher Intelligenz (KI) ist eine treibende Kraft der digitalen Transformation. Es ist daher notwendig, die Implikationen der Verwendung von KI in Bezug auf die IT-Sicherheit zu verstehen. Beide Gebiete zeigen zunehmende Überschneidungen und bilden technische Eckpfeiler für zukünftige Gesellschaften. Diesbezüglich existieren bereits Initiativen auf nationaler wie europäischer Ebene, welche die Schnittmenge von KI und IT-Sicherheit adressiern:

- In Frankreich wurde das wichtige Programm **"How to secure, certify and make reliable the systems involving AI?"**[8] initiiert. Dazu wird ein Programm Direktor (seit Anfang 2019 beim Dienst des Premierministers angesiedelt) Verantwortung für die Weiterentwicklungen in Bezug auf diese Herausforderungen tragen. Der thematische Fokus wird durch die Arbeiten in Zusammenhang mit dem "Villani-Report"[9] bestimmt.

- Das Bundesministerium für Bildung und Forschung veröffentlicht die Richtlinie "**Künstliche Intelligenz für IT-Sicherheit**".[10]

- SPARTA ist ein IT-Sicherheitsnetzwerk, an dem sowohl ANSSI als auch BSI beteiligt sind. Es ist Teil des EU-Programms Horizon 2020, um Spitzenforschung und -zusammenarbeit zu betreiben. Einen wesentlichen Eckpfeiler des SPARTA-Fahrplans stellt dabei das Programm SAFAIR dar, das Ansätze untersucht, Künstliche Intelligenz verlässlicher und mit höherer Resilienz zu nutzen, indem man die Erklärbarkeit erhöht und ein besseres Verständnis der Gefahren erhält[11].

- Während des 6. Deutsch-Französischen Forschungs-Kooperationsforums wurde KI, die die Privatsphäre (insbesondere bei Maschinellem Lernen) schützt, und verlässliche Architekturen als essentiell für die europäische Souveränität identifiziert[12]. Auch wenn dies nur indirekt ein Thema der IT-Sicherheit ist, gibt es doch wichtige Verbindungen, wenn man die Bezüge zu Biometrie und der kontinuierlichen Identifikation berücksichtigt. Bereits heute können KI-Techniken zur Umgehung einer Anonymisierung genutzt werden.

Eine Hauptfrage ist daher, auf welche Weise und mit welcher Wahrscheinlichkeit KI das Gleichgewicht zwischen Offensive und Defensive in der IT-Sicherheit beeinflussen wird. Auf der einen Seite arbeiten KI-Systeme nicht perfekt und ihre Schwächen können durch Angreifer ausgenutzt werden. Auf der anderen Seite stellt KI eine nie dagewesene Chance dar, eine Vielzahl innovativer Anwendungen zu schaffen.

Da der äußere Druck inzwischen sehr hoch ist KI-Anwendungen zu installieren, ist ein gutes Verständnis dieser Technik erforderlich, um eine entsprechende gemeinsame Stärkung der IT-Sicherheit zu erreichen. Der folgende Teil des gemeinsamen Lagebildes wird daher einige Beispiele beleuchten, die verschiedene Einflüsse der KI auf das aktuelle digitale Leben und die digitale Transformation haben. Zuvor soll allerdings ein kurzer Überblick gegeben werde, welche Elemente mit der Bezeichnung KI gemeint ist.

8 https://www.gouvernement.fr/sites/default/files/contenu/piece-jointe/2018/09/certification_ia.pdf (for the description of this challenge see page 3).

9 https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf

10 https://www.bmbf.de/foerderungen/bekanntmachung-2187.html

11 https://ssi.gouv.fr/uploads/2019/02/press-release-sparta.pdf

12 p. 38 of https://www.bmbf.de/upload_filestore/pub/BMBF_DF_FF_Dokumentation.pdf

KI

Die Bekanntheit von KI wurde durch die Erfolge im Machine Learning (ML - und hier insbesondere im Deep Learning (DL), Teilgebiete der KI) initiiert. Deep Learning wird durch Neuronale Netze implementiert, die viele Ebenen künstlicher Neuronen enthalten und mit enormen Datenmengen trainiert werden. Daher feierte DL Siegeszüge, seit große Datenmengen und (durch Graphical Processing Units – GPU) auch entsprechende Rechenleistungen verfügbar waren.

Zunächst erlaubte DL im Jahr 2012 in der Bilderkennung stark reduzierte Fehlerraten, wodurch menschliche Wiedererkennungsraten überschritten wurden[13]. Als zweites kam es im Jahr 2016 im Bereich der Computerspiele zum Sieg des Programms AlphaGo[14] gegen den Go-Meister Lee Sedol.

KI ist nicht auf das Thema Deep Learning begrenzt und hat eine Geschichte, die bis in die Pionierarbeiten der Kybernetik aus den 40er Jahren zurückreicht, aus der weiterhin Ansätze verfolgt werden. Einige basieren auf symbolischen Regeln, andere auf Maschinellem Lernen, das teilweise durch biologische Äquivalenzen inspiriert wurde. Dieser Hintergrund sollte bei der Einordnung der folgenden Ausführungen in den Gesamtkontext der KI berücksichtigt werden. Das Lagebild wird hauptsächlich Themen im Bereich des ML adressieren, da ML durch seine Abhängigkeit von großen Datenmengen für die IT-Sicherheit interessant wird.

Die Nutzung neuronaler Netze, die durch enorme Datenmengen trainiert werden, kann als Paradigmenwechsel gesehen werden. Das Wissen wird in diesen Netzen in einer Unzahl von Parametern (numerischen Werten) kodiert, die die statistische Struktur der Daten repräsentieren.

Dabei kann nur implizites Wissen generiert werden und es erweist sich als extrem schwierig, explizite Haupt-Determinanten für eine Entscheidung (z.B. durch Untersuchung der Parameter) zu identifizieren. Dies kann als fehlende Transparenz und Bedeutungsebene der Methoden interpretiert werden.

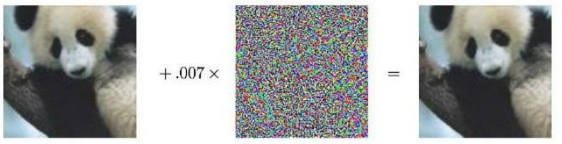
Trotz der großen Fähigkeit neuronaler Netze Generalisierungen zu erzielen, kann das Verhalten der NN in neuen seltenen Situationen nicht vorhergesehen werden. Auch wenn ihre Anwendung für einige Zeit im laufenden Betrieb getestet wurde, kann ihr erfolgreiches Arbeiten nur für den Bereich der bekannten Daten beobachtet werden. Dieser Bereich ist den Test- und Trainingsdaten sehr ähnlich. Dies kann als Verlässlichkeitsproblem interpretiert werden, da ein verlässliches KI-System mit großen Bereichen der möglichen Eingangsdaten und wechselnden Situationen arbeiten können muss.

Wie auch die Europäische Union[15] unterstreicht, sollte das Ziel einer "vertrauenswürdigen KI" verfolgt werden, die ausreichend Transparenz und Verlässlichkeit gewährt.

Schwachstellen der KI

- KI-Systeme sind anfällig für manipulative

Angriffe. Dies sind z.B. menschlich nicht wahrnehmbare Manipulationen legitimer Eingabedaten, die künstlich generiert wurden. Ein bekanntes Beispiel für diese Anfälligkeit ist der Pandabär, der mit hoher Verlässlichkeit als Gibbon erkannt wird, nachdem das ursprüngliche (richtig erkannte) Bild mit „manipuliertem Rauschen" überlagert wurde image[16]:



13 https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-networks

14 AlphaGo is based on a deep convolutional network to guide its exploration of the game, but also on Monte-Carlo methods and reinforcement learning.

15 Cf. The "Ethics Guidelines for Trustworthy AI" prepared by the High-Level Group on Artificial Intelligence set up by the European Commission https://bdi.eu/media/themenfelder/digitalisierung/publikationen/20190201_Stellungnahme_BDI_Draft_Ethics_Guidelines_for_Trustworthy_AI.pdf: it includes notably seven key requirements for the realization of Trustworthy AI among those technical robustness and safety and transparency.

16 Goodfellow IJ, Shlens J, Szegedy C (2014) Explaining and harnessing adversarial examples. arXiv. Available online at: https://arxiv.org/pdf/1412.6572.pdf).

Diese Angriffe wurden für eine Vielzahl von Eingangssignalen (Bild, Video, Ton)[17] demonstriert. Dabei bleibt das Prinzip immer gleich: Absichtlich gestörte Bilder (oder Audiodaten) wirken auf das Eingabespektrum des neuronalen Netzwerks, was zu einer Fehlklassifizierung des Inhalts oder einer angeblich zuverlässigen (aber falschen) Klassifizierung führt. Dies sollte zwar ungefährlich für eine Anwendung sein, die Bilder für eine Textsuche in einer großen Bilddatenbank beschriftet, da ein Fehler nur eine weniger nützliche Reihenfolge der Anzeige bedeutet. Sobald jedoch sicherheitskritische Anwendungen wie die autonome Fahrzeugkontrolle, die Grenzkontrolle oder die Analyse medizinischer Daten angesprochen werden, müssen diese qualitativ neuen Angriffsvektoren aufgrund des KI-Einsatzes berücksichtigt und geeignete Abwehrmechanismen für diese Angriffe entwickelt[18] und implementiert werden.

Schwächen von KI-Systemen, besonders qualitativ neue und KI-spezifische Angriffsvektoren erfordern einen sorgfältigen Umgang beim Ausrollen von Lösungen. Für kritische Sicherheitsanwendungen müssen effektive Verteidigungs- und Bewertungsstrategien entwickelt werden.

- Die Gefahr des **zufälligen Versagens** besteht für ML-Algorithmen darüber hinaus. So wurde ein Algorithmus trainiert, der Fotos von Huskies und Wölfen unterscheidet. Er schien zuverlässig zwischen den hundeähnlichen Tieren zu differenzieren. Doch es zeigte sich, dass dies mehr durch die Abbildung von Schnee im Hintergrund beeinflusst wurde als durch die Charakteristika der Tiere. Im Kontext der Informationssicherheit müssen noch weit weniger offensichtliche Korrelationen aufgedeckt werden.

Dies unterstreicht die Notwendigkeit weiterer Entwicklungen bei der Interpretierbarkeit von KI.

- Weit verbreitete ML-Methoden (insbes. NN) werden derzeit in großen Entwicklergemeinschaften kommuniziert. Nicht alle beteiligten Entwickler haben Zugriff auf Datenmengen, die große Unternehmen wie Google und Facebook zur Verfügung stehen. Um den Datenaustausch zu vermeiden und Schulungszeiten zu verkürzen, werden vortrainierte, allgemeingültige Modelle eingesetzt, die mit großen Mengen von Textdokumenten trainiert werden. Ein Nutzer des Modells muss anschließend nur mit einem kleinen Textkorpus (Menge der Dokumente) trainieren, um spezifische Ziele zu erreichen. Der Austausch dieser trainierten Neuronalen Netze - Transfer Learning (TL) - stellt eine Möglichkeit dar, implizites Wissen an Kunden weiterzugeben. Dieses übertragene Wissen kann, wie jede Art von Information, schadhaft manipuliert werden, was als "Data-Poisoning" (Vergiftung der Daten) bezeichnet wird. Dies kann von jedem in der Lieferkette, einschließlich des Herstellers der Datensätze, durchgeführt werden. Dieser könnte auch unbeabsichtigt interne Informationen preisgeben, die sich in den Daten verbergen. So ließen sich durch unbefugte Personen in den Daten verzerren. In diesem sich schnell entwickelnden Feld ist der Austausch von Wissen wertvoll, aber die freie Verfügbarkeit vieler Informationen auf Entwicklerplattformen (wie Github) hat negative Informationssicherheitsaspekte, da viele Teilnehmer in offene Lieferketten eingreifen könnten. Ein weiteres Szenario ist das kontinuierliche Training eines NN durch laufende Benutzerinteraktionen im Internet. Wenn die so trainierten NNe Transferdaten erzeugen, die von anderen ML-Prozessen verwendet werden, ist eine Manipulation der Eingabedaten z.B. mit geeigneten Chat-Bots möglich. Diese Roboter können wiederum als Angriffswerkzeug KI verwenden (siehe unten).

Offene Lieferketten und Transfer-Learning können misbraucht werden, um Daten zu "vergiften" (data poisoning) oder Modelle zu manipulieren.

- Die Verbreitung von KI-Systemen kann auch

17 It seems a priori more difficult to design adversarial examples adapted to data with other types of structure, like the ones encountered in classical cyber-security context (net traffic data, logs, etc.).

18 For a recent paper on robustification against adversarial attacks, see: "Robust Neural networks using Randomized Adversarial Training" (https://arxiv.org/abs/1903.10219). However, as robustification is mostly based on randomization, this can be to the detriment to the NN's accuracy.

neue Kanäle für Angriffe eröffnen. Wenn ML-Methoden als stark genug erachtet werden, um eine natürliche Interaktion über Kanäle wie Bild und Ton (Bild- und Spracherkennung) zu ermöglichen, könnte dies als vorherrschende Schnittstelle zur Steuerung von Computern oder IoT-Geräten (Bsp.: Hausautomation) eingerichtet werden. Während in der Vergangenheit Software aufgrund von Stapelüberlauf oder verborgener Funktionen ausnutzbar war, wirken sich Schwächen der neuen Kanäle allein auf der Informationsebene aus und sind schwerer zu erkennen und zu mildern. Bspw. besteht der "Delphinangriff"[19] im Senden von Sprachbefehlen mit Ultraschallwellen (mit Frequenzen > 20 kHz) für persönliche Assistenzsysteme wie Siri und Alexa. Diese Befehle können von IoT-Geräten wie Smart-TVs gesendet und von den Nutzern aufgrund der hohen Frequenzen nicht erkannt werden. Ähnliches könnte für optische Geräte im Infrarotbereich möglich sein. Wenn viele einzeln operable Systeme auf diese Weise zusammenarbeiten, können unvorhergesehene Effekte wie Voice Squatting auftreten. Dies sind phonetisch ähnliche Befehle, die von Drittanbietern hinzugefügt werden und orginäre Sprachbefehle überdecken.

Diese Schwachstellen sind nicht per se durch die Verwendung von KI begründet, zeigen aber die globale Vergrößerung der Angriffsfläche, wenn KI in die verteilte Steuerung von Informationssystemen integriert wird. Eine Gefahr ist das zweifelhafte Vertrauen durch breite Kopplungen.

KI als Schadsoftware (Waffe)

Aktuelle Anwendungen Maschinellen Lernens, die zwei Neuronale Netze als Gegenspieler verwenden, haben Bedrohungspotential.

- Es wurde gezeigt, dass Bilderkennung mit Hilfe von KI so getäuscht werden kann, dass sie präsentierte Daten nicht korrekt erkennt. Ein Trendthema in der KI-Entwicklung ist das **Generative Adversarial Networks (GAN)**. Obwohl der Name es vermuten lässt, sind GANs nicht für Angriffe konzipiert worden. Neuronale Netze werden dabei verwendet, um Schwächen eines Gegenspielers (ebenfalls ein NN) auszunutzen. Diese Technik kann, wie andere technische Instrumente auch, missbraucht werden, um z.B. Bilder zu finden, die implementierte Bilderkennungssysteme absichtlich täuschen. Auch künstliche Fingerabdrücke oder verschmolzene Bilder verschiedener Gesichter, die mehreren Benutzern als Authentifizierung dienen, können damit erzeugt werden. Obwohl GANs zur Ausnutzung von ML-Schwachstellen dienen, tragen sie nicht zur Erhöhung der Transparenz der Methoden bei, da es ihnen gleichfalls an Transparenz mangelt[20].

GANs sind ein gutes Beispiel für die duale inhärente Natur der meisten KI-Techniken.

- Eher klassische ML-Ansätze können ähnliche Ergebnisse wie GANs erzielen. Bekanntes Beispiel ist die Täuschung eines Gesichtserkennungsalgorithmus mit einer Brille mit geeignetem Muster auf der Fassung. In diesem speziellen Fall führt das Tragen der Brille zur Identifizierung als andere Person.[21]

KI kann KI angreifen und dies ist aktives Forschungsfeld, auch wenn die Forschung nicht dem Angriff dient.

- „Social Engineering“ ist als Angriffsvektor weit verbreitet. In Zukunft könnten KI-generierte Deep Fake-Videos hierbei eine stärkere Rolle spielen. Sie stellen bekannte und unbekannte Personen dar, um Handlungen zu bezeugen, die nie stattgefunden haben. Dabei wird in aufgenommenen Szenen das Gesicht der Zielperson auf das eines anderen Akteurs projiziert.

19 https://www.heise.de/forum/heise-online/News-Kommentare/Amazon-will-Alexa-das-unkontrollierte-Lachen-austreiben/DolphinAttack-gegen-Alexa/posting-31999888/show/

20 Karras, T.; Laine, S. & Aila, T. A Style-Based Generator Architecture for Generative Adversarial Networks.

21 Sharif, Bhagavatula, Bauer and Reiter: Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS ‘16). ACM, 2016

Dies könnte von Angreifern bei der Video-identifizierung oder zur Diskreditierung von Personen ausgenutzt werden. Auch wenn später gezeigt werden kann, dass das Videomaterial manipuliert wurde, kann es große Auswirkungen auf öffentliche Prozesse wie Wahlen haben, zur Diskreditierung öffentlicher Personen oder Erpressung von Lösegeld dienen.

Social Engineering, bei dem KI gefälschte Dokumente erzeugt, die dem Opfer den Anschein von Vertrauenswürdigkeit vermitteln, aber auch die durch KI unterstützte Desinformation sind gute Beispiele dafür, wie die Sicherheit häufig eher auf der semantischen als auf der technischen Ebene gefährdet ist.

- Ein weiterer Ansatz zur bösartigen Nutzung von KI besteht darin, Sicherheitsfunktionen wie HIP (Human Interaction Proof) zu umgehen. Diese nutzen komplexe Bilderkennungsaufgaben zur Verifizierung von menschlichen Interaktionen im Internet. Diese Art von Test wird bspw. zur Vermeidung automatischer Kontoerstellung für Internetanwendungen, die Personen als Akteure benötigen. Sie verwenden oft ein Bild von Zeichen, die verzerrt dargestellt werden, sodass eine gewöhnliche optische Zeichenerkennung (OCR) nicht funktioniert. Mit zunehmenden Fähigkeiten der KI wird diese Art des Schutzes schwächer. Ein anderes einfacheres Beispiel kann das Sortieren von veröffentlichten Passwortdatenbanken nach Nützlichkeit für ihre schädliche Anwendung sein. So wird z.B. das Kennwort "tre$or123“ eher für eine Bankanwendung als für ein Video-Streaming-Portal verwendet. Die Möglichkeiten einer solchen Verwendung von ML wird mit zunehmender Anzahl von Veröffentlichungen größer.

Klassische und weit verbreitete Sicherheits-mechanismen können durch KI geschwächt werden.

- KI in Verbindung mit großen Datenmengen und physikalischer Mustererkennung kann zur Deanonymisierung von Nutzern und Geräten, d.h. zur Kombinierung (Extrahierung) von Daten verschiedener Quellen und anschließenden Identifizierung verwendet werden. Der u.g. Seitenkanalangriff kann verwendet werden, um Hardware zur ver- und entschlüsseln.

- Oft bestehen bei Malware Schwierigkeiten darin, dass angegriffene Systeme eine große Vielfalt von veränderlichen Eigenschaften besitzen. Einige verfügen über Intrusion Detection-Systeme (IDS), andere über AV-Software und wieder andere unterstützen nicht alle üblichen Kommandos oder haben sehr restriktive Ausführungsrichtlinien. Um sich an diese diversen Umstände anzupassen, kann lernende Software diesen Anforderungen eher gerecht werden. Ein Rückkanal zum Entwickler der Schadsoftware kann Trainingsdaten für neue Versionen bereitstellen. Diese Daten speisen kein ML-System, aber KI ist mehr als reines Maschinelles Lernen. Sogar ausgefeilte Datenbanken, die logische Eingaben verarbeiten, sind Teil der KI. KI kann daher Autonomie fördern.

Schadsoftware könnte mit der Hilfe von KI unabhängiger von externer Unterstützung (C&C Server) werden, was ihre Entdeckung erschwert.

KI als Mittel zur Verteidigung

Auf der anderen Seite wurden verschiedene KI-Anwendungen eingeführt, die auch den Kampf gegen Cyberattacken unterstützen.

Kombiniert mit bestehenden Ansätzen haben KI-Methoden in der Cybersicherheit das Potenzial, die Abwehr in verschiedenen Phasen zu verbessern: Bei der Entwicklung und Bewertung von Produkten, bei der Erkennung von Angriffen und (mit Entscheidungs-hilfen) in der Phase der Reaktion nach Angriffen.

Aufgrund der großen Datenmenge, mit der Cyber-Analysten konfrontiert sind, und der wachsenden Komplexität von Angriffen könnten sich KI-Methoden als sehr nützlich erweisen, indem sie die Arbeit der Verteidiger teilweise automatisieren.

- Moderne AV-Software erkennt Malware nicht nur anhand von Signaturen, sondern auch durch Algorithmen des Maschinelles Lernens, die mit den Eigenschaften tausender bekannter Schadprogramme trainiert wurden. Viele Parameter der Software können als Indikator dienen, sogar die Struktur des Codes. Daher müssen Bedrohungsakteure innovativ sein und ihren Code ändern, um eine Erkennung zu vermeiden.

KI hilft bei der Erkennung von Schadsoftware.

■ Bei der Beobachtung des Netzverkehrs entstehen große Datenmengen. Diese liefern für Angriffe charakteristische Anomalien. Obwohl die Anwendung von KI nicht leicht nachzuweisen ist, arbeiten viele Institutionen daran. Dies wird angewendet im lokalen (Unternehmen, Organisation) oder globalen Kontext (Internet Service Provider (ISPs)) sowie in Content Delivery Networks (CDN) angewendet werden.

KI findet Anomalien im Netzverkehr.

■ Eine der ältesten Anwendungen von KI ist die Detektion von Spam-E-Mails. Während das Auffinden von Schlüsselwörtern einst reichte um Spam-E-Mails zu identifizieren, sind Spam-Wellen mit gefährlichen Anhängen oder URLs professioneller. Moderne Spam-Filter können diese identifizieren, solange die Algorithmen kontinuierlich trainiert werden.

KI verhindert bestimmte Angriffsvektoren.

■ KI kann bei der Angriffsdetektion auf biometrische Identifikationssysteme wie Gesichts- oder Fingerabdruckerkennung verwendet werden. Eine Herausforderung sind verschmolzene Bildern von Gesichtern. Die Verwendung von KI mit ausreichenden Trainings- und Testdaten führt jedoch zur robusten Erkennungen sogenannter Morphing-Attacken.

Durch KI werden Betrugsversuche aufgedeckt.

■ Moderne kryptografische Hardware nutzt komplexe Algorithmen, die sich bei der Beobachtung von Seitenkanälen nur schwer untersuchen lassen. Dennoch wurden KI-Methoden erfolgreich eingesetzt, um Schwächen solcher Geräte nachzuweisen. Dabei verwendet der Analyst Informationen wie Stromverbrauch oder Ausführungszeit, um den geheimen Schlüssel des Geräts zu extrahieren. Klassische Methoden neigen hierbei zu Fehlern durch Ungenauigkeiten in den Modellannahmen wie Verzögerungen und Rauschverteilung. Seitenkanalanalysen, die auf ML-Techniken basieren, reagieren weniger empfindlich. Sie sind robuster und benötigen weniger Vorverarbeitung. Als Waffen verwendete KI-Werkzeuge (siehe oben) können helfen, Software und Hardware auf ihre Sicherheit hin zu untersuchen.

KI hilft Hard- und Softwarekomponenten zu untersuchen und zu härten.

■ Schließlich kann das enorme Informationsangebot im Netz genutzt werden, um die aktuelle IT-Sicherheit zu bewerten. Momentan verzeichnet die automatische Interpretation von Texten durch ML-Algorithmen Forstschritte. Daher ist die automatische Extraktion von Entitäten und die automatische Anreicherung einer Dokumentensammlung möglich. Dies ist Gegenstand aktueller Forschung.

Das Bewusstsein für Gefahren lässt sich durch automatische Textverarbeitung via KI verbessern.

Viele zukünftige KI-Anwendungsfälle sind denkbar, z.B. kontinuierliche Authentifizierung von Nutzerverhalten, die Erkennung neuer Datenlecks durch die Beobachtung von Login-Frequenzen oder die Vorhersage von zukünftigen Angriffswellen.

Es gibt viele zukünftige KI-Anwendungen.

Zusammenfassung

In Bezug auf KI haben beide Behörden Erfahrung sich Herausforderungen und Möglichkeiten in Bezug auf die Cyber-Sicherheit zu stellen. ANSSI und BSI sind gemeinsam, sowohl direkt als auch indirekt, in KI-Projekten auf europäischer Ebene beteiligt, so z.B. bei SPARTA, das Teil des strategischen Forschungs- und Entwicklungsprojekts Hprizon2020 ist. Dieses ist nur eines von zahlreichen Projekten, das von beiden Behörden verfolgt wird, um ihre nahe Zusammenarbeit sowohl auf der Forschungs- und Entwicklungsebene als auch auf dem technisch-operativen Level zu erweitern.

Im Hinblick auf die breit aufgestellte und tief gehende bilaterale, wenn auch nicht ausschließliche, Deutsch-französische Zusammenarbeit dient das Deutsch-französische IT-Sicherheitslagebild zur Erhöhung und deutlichen Unterstreichung der Sensibilität der Leserschaft für aktuelle Themen der Cyber-Sicherheit, um die Cyber-Sicherheitslage langfristig zu verbessern.

Agence nationale de la sécurité
des systèmes d'information
51, boulevard de La Tour-Maubourg
75700 Paris 07 SP

Bundesamt für Sicherheit in
der Informationstechnik
Postfach 200363
D 53133 Bonn