Sultanow, Eldar; Bauckhage, Christian; Knopf, Christian; Piatkowski, Nico

Sicherheit von Quantum Machine Learning

Wirtschaftsinformatik & Management

Provided in Cooperation with:
Springer Nature

Suggested Citation: Sultanow, Eldar; Bauckhage, Christian; Knopf, Christian; Piatkowski, Nico (2022) : Sicherheit von Quantum Machine Learning, Wirtschaftsinformatik & Management, ISSN 1867-5913, Springer Fachmedien Wiesbaden GmbH, Wiesbaden, Vol. 14, Iss. 2, pp. 144-152, https://doi.org/10.1365/s35764-022-00395-6

This Version is available at:
https://hdl.handle.net/10419/313215

Sicherheit von Quantum Machine Learning

Cyberkriminalität bewegt laut Cybersecurity Ventures weltweit schon heute das meiste Geld [1]. Werden Quantencomputer noch dazu beitragen oder die IT-Sicherheit erhöhen? Sie bieten neue Angriffsflächen und können „klassische" Sicherheitsmechanismen brechen, aber auch die Verteidigung optimieren. Maschinelles Lernen (ML) wird dabei als Quantum Machine Learning (QML) eine wichtige Rolle spielen.

Eldar Sultanow, Christian Bauckhage, Christian Knopf und Nico Piatkowski

Wirtschaftsinformatik & Management 2022 • 14 (2): 144–152
https://doi.org/10.1365/s35764-022-00395-6
Angenommen: 4. Januar 2022
Online publiziert: 24. März 2022

Klassisches Machine Learning steht unter dem Begriff „Neural Cryptography" bereits heute im Dienste der Kryptographie. Etwa lassen sich sogenannte *Tree Parity Machines* (spezielle neuronale Netze) als Public-Key-Kryptosystem einsetzen, das den Schlüsselaustausch zwischen zwei Parteien vollzieht und damit eine Alternative zum Diffie-Hellman-Schlüsselaustauschprotokoll darstellt [2]. Das Prinzip dieses neuronalen Schlüsselaustauschs ist vergleichbar mit der Synchronisierung zweier chaotischer Oszillatoren in der Chaoskommunikation. Ein weiterer Anwendungsfall ist *Mutual Learning*, das zwei kommunizierenden Parteien ermöglicht, gemeinsam einen geheimen Schlüssel über einen öffentlichen Kanal zu erstellen. Dabei kann ein Angreifer nur einseitig lernen und hat praktisch keine Möglichkeit denselben geheimen Schlüssel zu erstellen, den die beiden Parteien nutzen [2]. Die sogenannte Quantenschlüsselverteilung (Quantum Key Distribution, kurz QKD) ermöglicht zwei entfernten Parteien, geheime Schlüssel in einer nicht vertrauenswürdigen Umgebung auszutauschen, ohne von Dritten abgehört zu werden. Die Leistung und praktische Sicherheit von QKD-Systemen optimieren Liu et al. [3] durch maschinelles Lernen.

Auch positionsbasierte Kryptographie (PBC) nutzt klassisches Machine Learning. Sie dient zur Verifikation, dass ein Adressat sich an (oder in der näheren Umgebung) eines bestimmten Ortes befindet [4]: Wenn wir bei einem Kassierer Geld auf unser Konto einzahlen wollen, ist ganz ähnlich die Frage entscheidend, ob er in einer Bank arbeitet oder in einem Supermarkt. Ein weiterer Anwendungsfall ist die automatische Erhebung von Straßenbenutzungsgebühren. Die klassische PBC stößt in der Positionsverifikation an ihre Grenzen gegen Angreifer, die alle Positionen außer der behaupteten Position des Verifizierers (prüfenden Anwenders) kontrollieren. Dies führte zu einem neuen Gebiet, nämlich der positionsbasierten Quantenkryptographie (PBQC), die einen solchen Angriff (die Kontrolle aller Positionen außer der behaupteten Position des Verifizierers) verhindern soll. Christian Schaffner [5] gibt die Geschichte der positionsbasierten Quantenkryptographie anhand relevanter Publikationen wieder. Lau und Lo [4] sowie Beigi und König [5] schlagen eine Härtung von positionsbasierten Quantenkryptographieprotokollen gegen Verschränkungsangriffe vor. Bykovsky [6] setzt neuronale Netze in der positionsbasierten Kryptographie für vernetzte mobile Roboter ein und zeigt, wie sich die Vielfalt vorhandener Verifikationsverfahren erweitern lässt.

Weitere Anwendungsfälle für (Q)ML sind die Klassifizierung verschlüsselten Netzwerkverkehrs (beziehungsweise verschlüsselter Daten allgemein), die Bekämpfung von Steganographie, die homomorphe Verschlüsselung und die symmetrische Verschlüsselung basierend auf *Counter Propagation Networks* [2]. Alani liefert hier auch Anwendungsbeispiele für die Kryptoanalyse: Beispielsweise wird ML eingesetzt bei Seitenkanalangriffen oder von Angreifern für Known-Plaintext-Attacken mittels neuronaler Netze [7]. Hier wird ein Chiffrat entschlüsselt, indem Plaintext-Ciphertext-Paare analysiert werden. Zum Einsatz kommen kann ML auch für Angriffe zur Schlüsselermittlung, zur Ermittlung von Benutzeraktionen in verschlüsseltem

Dr. Eldar Sultanow [1, 3] (✉)
ist als Enterprise Architect bei Capgemini tätig und forscht an der Universität Potsdam.
eldar.sultanow@wi.uni-potsdam.de

Prof. Dr. Christian Bauckhage [2]
ist Lead Scientist für Machine Learning am Fraunhofer IAIS und Professor an der Universität Bonn.
christian.bauckhage@cs.uni-bonn.de

Christian Knopf
ist Cybersecurity Defense Advisor bei Capgemini.

[1] *Universität Potsdam, Potsdam, Deutschland*

[2] *Bonn, Deutschland*

[3] *Universitätsgesellschaft Potsdam e.V., Potsdam, Deutschland*

Dr. Nico Piatkowski
ist Senior Scientist am Fraunhofer IAIS.

Netzwerkverkehr oder etwa für das Verhindern ML-basierter Angriffe in physisch nicht klonbaren Funktionen (PUF).

Angriffe auf klassische Kryptographie durch Quantencomputer

Die Quanteninformatik wird die IT-Sicherheit global dauerhaft und tiefgreifend verändern, da sie für bestimmte mathematische Problemklassen einen exponentiellen Vorteil gegenüber der klassischen Informatik bietet. Eine Anwendung von Quantum Computing hat bisher am meisten Aufmerksamkeit bekommen: Heutige Public-Key-Kryptographie kann mit Quantencomputern effizient angegriffen und gebrochen werden. So werden die Verschlüsselungsverfahren statt in Jahren in Minuten umkehrbar sein.

Public-Key-Kryptographie wird eingesetzt, wenn Daten zwischen zwei Systemen oder Netzen verschlüsselt werden – also realistisch überall dort, wo Daten ausgetauscht werden. Insbesondere im HTTPS-Protokoll und in der Zertifikatsstruktur (etwa bei der Signierung von Software, VPN-Tunneln oder E-Mails) sind diese Verfahren sehr wichtig. Die Basis für Angriffe auf Public-Key-Kryptographie ist der Shor-Algorithmus [8]. Er wurde bereits 1994 konzipiert und hat einen der wichtigsten Beiträge seiner Zeit zum Quantum Computing geleistet: Er hat erstmals die Frage eindeutig beantwortet, ob Quantencomputer *überhaupt* Vorteile gegenüber klassischen Berechnungsmethoden haben. Der Shor-Algorithmus reduziert den Zeitaufwand zur Faktorisierung großer Zahlen von subexponentieller Zeit (größer als jedes Polynom) mit klassischen Computern auf polynomielle Zeit mit Quantencomputern. In Zukunft könnten beispielsweise Geheimdienste HTTPS, SSL, VPNs und jeden anderen verschlüsselten Datenverkehr, jede verschlüsselte Datei und jede verschlüsselte Hardware in einen Quanten-Decryptor einspeisen [9]. Ähnlich sieht SAP die Unbezwingbarkeit der Blockchain und somit die Sicherheit von Kryptowährungen bedroht [10].

In der Quintessenz müssen, um das Kräfteverhältnis zwischen Cyberangriff und -abwehr wieder zugunsten der Sicherheit zu verschieben, neue kryptographische Methoden (aus der Quanteninformatik) Einzug halten. Dies geschieht bereits – das Stichwort ist Quantenkryptographie. Dabei handelt es sich keineswegs erst um theoretische Konstrukte, sondern um bereits realisierte praktische Anwendungen. So haben 2021 zwei Bundesbehörden in Bonn mittels Quantenschlüsselverteilung eine abhörsichere Videokonferenz über Glasfaser und Freistrahlkanäle erfolgreich durchgeführt [11]. Zwar ist die praktische Sicherheit von QKD kompromittierbar (ein typisches Beispiel ist der sogenannte „Quantum Man-in-the-Middle-Angriff" auf den Kalibrierungsprozess [13]), jedoch kann klassisches ML diese Angriffe abwehren. So stellen Guo, Yang und He [12] gegen Angriffe auf die Quantenschlüsselverteilungsart CVQKD einen Verteidigungsmechanismus vor, der auf einem neuronalen Netz basiert. Al-Mohammed et al. [13] beschreiben ebenfalls ein ML-basiertes Vorgehen, mit dem sie bei Angriffen auf QKD im Anwendungskontext von 5G-IoT-Netzwerken die Angreifer in 99 % der Fälle identifiziert haben.

Die Zeit ist also reif, sich mit Auswirkungen auf die eingesetzten kryptographischen Primitive und Standards zu befassen. Dem kommt seit einiger Zeit die Post Quantum Cryptography Initiative des NIST in den USA nach [14].

Limitierte Leistungsfähigkeit heutiger Quantencomputer

Heute existierende Quantencomputer stehen erst an der Schwelle, in bestimmten QML-Anwendungen mehr zu leisten, als es Höchstleistungsrechner können. Erste Erfolgsbeispiele aber gibt es – etwa bei der Berechnung sich überlappender Wellenfunktionen, um die Eigenschaften eines Quantensystems zu eruieren bzw. zu überprüfen. In einem Experiment wurde eine Schrödingergleichung mit 2^{53} Basiszuständen mit dem Sycamore-Quantencomputer dargestellt. Binnen dreieinhalb Minuten erzeugte der Quantencomputer eine Million Zahlenreihen (aus je 53 Bits) als Ergebnis. Diese Ergebnisse können statistisch analysiert werden, um Verteilungen zu ermitteln. Ein Cluster aus Hochperformance-Computern mit einer Million Prozessoren würde circa 10.000 Jahre für die Simulation derselben Gleichung benötigen [15, 16].

Trotz der derzeit begrenzten Einsatzmöglichkeiten ist QML ein vielversprechendes Feld, und es findet entsprechend viel Forschung an Algorithmen und Modellen statt. Wissenschaftliche Manuskripte weisen einen Geschwindigkeits- und Genauigkeitsvorteil von QML auf in naher Zukunft verfügbaren Geräten (sogenannten NISQ-Technologie/Noisy Intermediate-Scale Quantum Devices) gegenüber klassischem ML theoretisch nach [17–19]. Gleichermaßen nehmen wissenschaftliche Studien mit praktischem QML-Bezug zu. So stellen Senekane, Mafu und Taele [20] einen Ansatz für datenschutzkonformes QML vor, den sie an einem sensiblen Datensatz mit *Features* und *Target Labels* in der Brustkrebsdiagnostik erfolgreich verifiziert haben. Dixit et al. benutzen QML zur Erkennung von Cyberangriffen, indem sie Anomalien in Netzwerkaktivitäten identifizieren.

Zusammenfassung

- Quantum Machine Learning ist ein junges Gebiet der Quanteninformatik und mittlerweile praktisch implementierbar.
- Mit den zur Verfügung stehenden Quantencomputern kann QML seine Vorteile gegenüber klassischem ML noch nicht ausspielen.
- Das wird sich in den nächsten Jahren ändern.

Machine Learning auf Quantencomputern

Quantencomputer sind gut für das Lösen von Optimierungsaufgaben geeignet, bei dem Tensoren für Machine Learning zum Einsatz kommen. Je nach Verfahren können sie auch hier ihre Überlegenheit gegenüber klassischen Computern beweisen. So stellt Google etwa mit *TensorFlow Quantum* eine Bibliothek für hybrides quantenklassisches maschinelles Lernen bereit [23]. Weitere Frameworks sind PennyLane und QTN-VQC [24].

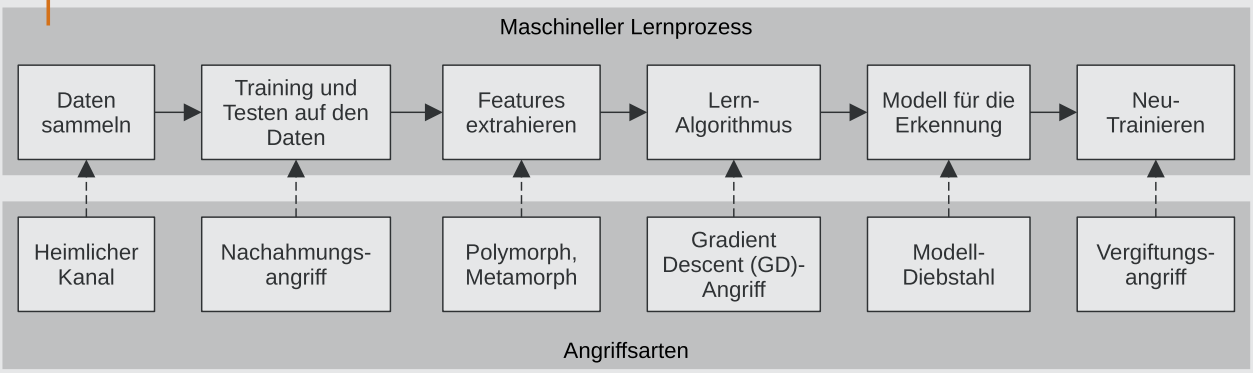
Lebenszyklus von Machine-Learning-Modellen

Machine Learning hat seine Hauptanwendung in der Klassifizierung – beispielsweise ob eine Mail in Spam oder Ham eingeordnet wird, ein Bild ein Verkehrsschild zeigt (und welches), welche Dokumente aus einem großen Satz als relevant eingestuft werden oder ob eine Datei unerwünschte Software enthält. Komplexere Modelle können etwa dazu verwendet werden, aus Eingabedaten Bilder zu erzeugen und Umgangssprache in korrekte Texte oder eine andere Sprache zu übersetzen. Das GPT-3-Modell hat 2021 als bisher komplexestes zum Schreiben von Texten für Aufsehen gesorgt [21].

Das generelle Vorgehen bei der Umsetzung maschineller Lernverfahren in konkreten Anwendungsszenarien lässt sich in verschiedene Phasen gliedern. Eine standardisierte Herangehensweise bietet CRISP-DM [22]. Hier wollen wir, losgelöst von Methoden wie CRISP-DM, kurz auf die grundlegenden Schritte eingehen.

Vor der Anwendung von ML-Verfahren steht die Sammlung von Daten. Hierbei ist es wichtig, dass die Daten stets den gesamten zu modellierenden Prozess abdecken. Darüber hinaus sollen die einzelnen Werte eines Datums, beispiels-

Abb. 1 Angriffsflächen auf Machine Learning im Lernprozess nach Wang et al. [23]



weise die mittlere Temperatur in einem Schmelzofen oder die Güte eines produzierten Werkstücks, stets mit gleichbleibender Präzision gemessen werden. Die Qualität der Daten entscheidet über den Erfolg des Modells nach der Lernphase.

Im zweiten Schritt, der Vorverarbeitung, werden die Daten in ein verarbeitungsfähiges Format überführt. So können die Methoden des maschinellen Lernens – mathematische Modelle – die Daten in Form von Matrizen oder Vektoren verarbeiten.

Die Gesamtheit der vorhandenen Daten wird in mindestens zwei Mengen aufgeteilt: die Trainings- und die Testmenge. Anhand der Trainingsmenge findet das eigentliche Lernen statt. Dabei werden mittels Methoden der numerischen Optimierung die Parameter eines ML-Modells bestimmt. Die Daten der Testmenge dienen dazu, die Genauigkeit des gelernten Modells zu bewerten. Es ist wichtig, dass die Trainings- und die Testmenge keine Schnittmenge haben; andernfalls würde ein Test lediglich überprüfen, ob die Trainingsdaten korrekt auswendig gelernt wurden.

Die Qualität des Modells zeigt sich im Umgang mit unbekannten Daten, in der Generalisierung. Ist die Generalisierungsfähigkeit hoch genug, so kann das Modell in die Anwendung überführt werden. Da in den meisten Prozessen ständig neue Daten anfallen, können diese genutzt werden, um das aktuelle Modell zu verbessern. Dies erfolgt entweder fortlaufend im sogenannten Datenstrommodell oder aber das alte Modell wird durch ein neues ersetzt, sobald genügend neue Daten vorhanden sind.

Ist das Modell trainiert und seine Qualität bestätigt, kann es im produktiven Einsatz seinen Zweck erfüllen. Verwendet wird ein Modell normalerweise so lange, bis es am Ende seines Lebenszyklus durch ein neues, mächtigeres Modell ersetzt wird.

Sicherheit spielt nicht erst im produktiven Einsatz beziehungsweise während der Anwendung des Machine-Learning-Modells eine wichtige Rolle, denn sie hängt – parallel zur Softwareentwicklung – von der Absicherung jeder der beschriebenen Phasen ab. Durch Einflussnahme auf die Datenkollektion oder das Training kann das resultierende Modell grundlegend kompromittiert und nutzlos gemacht werden.

Angriffe auf Machine-Learning-Systeme

Der maschinelle Lernprozess bietet Angriffsflächen für unterschiedlichste Arten von Attacken. Wang et al. stellen in **Abb. 1** für jeden der sechs beschriebenen Schritte des maschinellen Lernens einen möglichen Angriffspunkt dar – ohne Anspruch auf Vollständigkeit [23].

Kernthesen

- Noch geht von QML keine Gefahr für klassische IT-Sicherheitsmethoden und -systeme aus.
- Quanteninformatik allgemein (nicht QML) stellt bereits heute eine Gefahr für die klassische IT-Sicherheit dar (Angreifbarkeit asymmetrischer Verfahren).
- Quanteninformatik allgemein liefert bereits durch Quantenverschlüsselung einen Beitrag zu noch höherer IT-Sicherheit.

Eine Kategorisierung von Angriffen kann anhand dreier fundamentaler Kriterien erfolgen: Einflussmöglichkeiten des Angreifers (1), angegriffene Eigenschaft des Systems (2) und Angriffsspezifizität (3), siehe [23–24].

1. *Die Einflussmöglichkeiten des Angreifers* umfassen alle externen Größen, die auf den Lernprozess einwirken und so potenziell eine Einflussnahme gestatten. Bei einem Angriff auf Trainingsgrundlagen (Causative Attack) hat der Angreifer die Möglichkeit, die Menge der Trainingsdaten zu verändern. Da die Daten die Güte des Modells bedingen, kann das Lernergebnis auf diese Weise stark beeinflusst werden, auch wenn vielleicht nur ein kleiner Teil der Trainingsdaten verändert oder hinzugefügt wird. Das ist insbesondere bei extremen Inputdaten der Fall.
   Es besteht auch die Möglichkeit, nicht das ML-Modell an sich zu manipulieren, sondern Schwachstellen herauszufinden und auszunutzen. Beim Entdeckungsangriff (Exploratory Attack) werden durch Informationen über das Modell insbesondere Eingaben genutzt, die zur Fehlklassifizierung führen.
2. *Das zweite Kategorisierungskriterium ist die angegriffene Eigenschaft des Systems*. Es hat die auch sonst in der Cybersecurity verwendeten Ausprägungen: Integrität, Verfügbarkeit, Vertraulichkeit. Der Angriff auf die Integrität ist eine gezielte Fehlklassifikation von einzelnen Angriffsdaten. Der Angriff auf die Verfügbarkeit macht das Modell unbrauchbar, indem durch den Angriff viele Fehlklassifikationen vom ML-System ausgehen (falsch-positiv und/oder falsch-negativ). Beim Angriff auf die Vertraulichkeit kann der Angreifer Informationen über das Modell selbst oder über das Trainingsdatenset herausfinden – also zum Beispiel eine Form des Reverse Engineerings. Dies kann

als Grundlage zu weiteren Angriffen dienen: Nachdem zuerst das Modell nachgebaut wird, kann es zur Generierung bzw. Validierung von Angriffsdaten verwendet werden. So kann es beispielsweise Bilder generieren, die vom System falsch klassifiziert werden.

3. *Angriffsspezifizität* bezeichnet, wie zielgerichtet oder unspezifisch der Angriff erfolgt. Zielgerichtet bedeutet, dass der Angreifer einen klar umrissenen Effekt erzielen möchte – beispielsweise die Fehlklassifikation eines bestimmten Inputs. Unspezifisch dagegen bedeutet, dass der Angreifer mehr Flexibilität hat und nur irgendeinen Input kennen möchte, der vom System falsch klassifiziert wird.

Konkrete Beispiele für Angriffe auf Machine-Learning-Systeme

In der Praxis sind vielfältige Angriffe auf ML-Systeme zu finden, etwa zur Umgehung von Spamerkennung. Sie lassen sich unterschiedlichen ML-Ebenen bzw. den beschriebenen Kategorien zuordnen [25].

Modelldiebstahl. (Model Reconstruction) ist ein Angriff, bei dem Angreifer das gesamte ML-Modell in Erfahrung bringen und nachbauen (Reverse Engineering). Sie nutzen dazu das Wissen um die Trainingsdatenzugehörigkeit, können feststellen, ob bestimmte Eingaben Teil des Trainingsdatensets waren und ggf. den gesamten Satz der Trainingsdaten wiederherstellen [26, 27].

Adversarial Attacks. sind Angriffe in Brute-Force- bzw. Trial-and-Error-Manier, die das Machine-Learning-System mit unzähligen Eingabedaten befragen und dessen Reaktion beobachten. Dies wird so lange durchgeführt, bis eine vom System falsch klassifizierte Eingabe gefunden wird (unspezifischer Angriff) oder sogar zu einer vom Angreifer gewünschten Klassifizierung führt. Spamfilter können auf diese Art angegriffen werden, indem anhand der Beobachtung der Klassifizierungsergebnisse der Inhalt einer Spam-E-Mail so lange variiert und gesendet wird, bis der Filter ihn nicht mehr als Spam erkennt [25]. Als Verfeinerung dieses Angriffs lassen sich anhand der Reaktion die Eingabedaten verändern und anpassen. Im Ergebnis kann der Angreifer Teile der Eingabe selbst definieren und gleichzeitig eine Fehlklassifikation (unspezifischer Angriff) oder sogar eine von ihm gewünschte Klassifizierung erzielen. Angriffe auf Antivirenprodukte sind für Angreifer effektive Einfallstore – trotz bösartigem Code werden die Daten dann als unproblematisch eingestuft.

Vergiftete Daten. dienen zur Modellverzerrung: Indem der Angreifer Trainingsdaten verändert (bzw. zusätzliche Elemente hinzugefügt), kann er das resultierende ML-Modell zielgerichtet manipulieren. Die Klassifizierung lässt sich so insgesamt in eine bestimmte Richtung verlagern oder nur für bestimmte Elemente verschieben. Angreifer setzen diese Methode insbesondere bei selbstlernenden Systemen (Incremental Machine Learning) ein, wenn also einem Machine-Learning-Modell im Betrieb laufend neue Daten gegeben werden können, die das Modell verändern. Daten, die eigentlich in Kategorie A fallen, können dann immer weiter in Richtung B tendieren.

Diese Angriffsart kann prinzipiell alle inkrementellen ML-Systeme treffen. Im Falle eines Spamfilters [25] sollten als Spam klassifizierte, aber immer weiter in Richtung valide E-Mails tendierende Eingabedaten den Spamfilter insgesamt dazu bringen, zu viele Mails als Spam zu klassifizieren.

Bei sogenannten Hintertürangriffen modifiziert der Angreifer das ML-System so, dass es im Einsatz unter normalen Bedingungen korrekte Ergebnisse liefert, besondere Eingaben (beispielsweise unter Anwendung eines Geheimnisses, wie einzelnen Pixeln eines Bildes) jedoch zu vom Angreifer definierten Resultaten führen.

Angreifbarkeit von Quantum Machine Learning

Die im Bereich des klassischen ML betrachtete Methode der Adversarial Attack spielt auch im Kontext von QML eine große Rolle.

Lu, Duan und Deng [28] übertragen Schritt für Schritt das Adversarial Machine Learning in die Quantenwelt. Damit können sie Klassifizierer auf Quantenrechnern betrachten und so die ML-Angriffsflächen analysieren. Sie stellen Methoden vor, um Perturbationen für Quantum-Klassifizierer zu generieren – etwa Veränderungen von Inputdaten, die zu Fehlklassifizierung führen. Liu und Wittek [29] fügen in derselben Stoßrichtung eine Betrachtung des Aufwands von Verteidigungsmaßnahmen hinzu. Sie kommen zu dem Schluss, dass diese Schutzmaßnahmen zusätzliche Ressourcen binden, womit sie den Vorteil des Quantum Computing wieder aufheben. Liao et al. [30] schauen sich die Robustheit der Quantenklassifizierungen an. Sie kommen hier zu dem Schluss, dass die QML-Modelle unter gewissen Bedingungen robust genug und für den Einsatz unter realen Bedingungen geeignet sind.

Wiebe und Kumar [31] geben erste Einblicke, wie man unterschiedliche QML-Modellklassen gegen Quantum Adversarial Attacks härten kann.

Findet QML an einem entfernten Ort statt, ist die Frage nach einem Angriff, durch den externe Dritte unberechtigt

Handlungsempfehlungen

- Auseinandersetzung mit den Möglichkeiten von Quanteninformatik im praktischen Sinne, etwa durch Nutzung der zur Verfügung stehenden Frameworks
- Vorbereitung auf die nächste Ära: Antizipieren und Verproben der Probleme, die sich quanteninformatisch formulieren und umsetzen lassen

am Lernprozess teilnehmen oder diesen gar stören können, wichtig. Unter dieser Prämisse entwickeln Bang, Lee und Jeong [32] ein Protokoll für sicheres QML, das verteilt stattfindet.

Insgesamt muss gesagt werden, dass generalisierende Aussagen über die Sicherheit sowie die Angreifbarkeit von QML derzeit nur sehr schwer zu treffen sind. Die Forschung betrachtet größtenteils erst einzelne Verfahren oder spezielle Bedingungen. Dies liegt auch darin begründet, dass die Arbeiten an der Sicherheit von QML bislang ausschließlich theoretischer Natur sind, denn es gibt schlicht noch kein QML im Realwelteinsatz. Es ist allerdings davon auszugehen, dass die aus der klassischen ML bekannten Angriffsformen aus dem Forschungsfeld der Adversarial Attacks die größte Bedrohung auch für zukünftiges QML darstellen, sofern Zugriff auf das Modell besteht.

Fazit und Ausblick

Die Kryptoanalyse – das Brechen von Public-Key-Kryptosystemen beziehungsweise die Faktorisierung großer Zahlen mit zwei großen Primfaktoren – ist der am stärksten erwartete Anwendungsfall von Quantenrechnern. Hier wissen wir, dass Quantum Computing definitiv eine Bedrohung für die klassische IT-Sicherheit darstellt.

Sowohl bei hoher Komplexität als auch hinsichtlich ihrer Genauigkeit schlägt jede Grafikkarte beziehungsweise GPU einen heutigen Quantencomputer. Diese sind tonnenschwer, sehr aufwendig zu kühlen und von Umgebungseinflüssen abzuschirmen. Das macht sie teuer im Bau sowie im Unterhalt. Zudem stellen die Inputdaten für maschinelle Lernaufgaben auf Quantencomputern noch eine Hürde dar, die für klassische Hardware mit leicht verfügbarem, großem Speicher nicht besteht. Beides gilt für eine einfache Regression gleichermaßen wie für komplexe, tiefe neuronale Netze.

Machine Learning auf Quantencomputern zu betreiben, kann dennoch naheliegend sein. Im Gegensatz zur Kryptoanalyse wird QML insbesondere auf den bald schon verfügbaren NISQ-Devices möglich sein. Daher werden wir mit steigender Leistungsfähigkeit der Quantencomputer auch mehr und mehr Erfolge im Bereich QML wahrnehmen können.

Quantum Computing wird aus den oben genannten Gründen zunächst als Quantum Cloud oder Quantum as a Service (QaaS) zu Verfügung stehen – beziehungsweise befindet sich bereits in minimaler Form im produktiven Einsatz [33]. Inwieweit es sich für Unternehmen lohnen wird, eigene Quantum-Computing-Kapazitäten aufzubauen, wird sich zeigen. Die Entwicklung steht erst ganz am Anfang – vergleichbar mit dem Status von digitalen Computern in den 40er- und 50er-Jahren des letzten Jahrhunderts.

Langfristig gesehen ist das Potenzial von QML immens. Von Beginn an sollten in jedem Einsatz von Quantum Computing Sicherheitsaspekte eine hohe Priorität erhalten. Kann die Sicherheit der einzusetzenden QML-Verfahren, -Modelle und -Eingabedaten nicht umfassend gewährleistet werden, sollte zumindest eine entsprechende Risikoabschätzung durchgeführt werden.

Durch QML und den Zugriff auf entsprechende Quantenhardware werden neuartige Angriffe möglich. So kann ein Angreifer mit Zugriff auf ein QML-System ein klassisches ML-System viel schneller, effizienter und umfassender analysieren und angreifen. Weitere Verfahren aus dem nahen ML-Umfeld, etwa genetische beziehungsweise populationsorientierte (Brute-Force-)Algorithmen, werden auf die Quanteninformatik übertragen werden.

Aktuell stecken die theoretische QML-Forschung sowie die praktische Verfügbarkeit von Quantencomputern noch in den Kinderschuhen. Ein konkreter Einsatz von QML – und damit konkrete Angriffe auf und seitens QML – sind daher noch Zukunftsmusik. Dieser Stand wird sich voraussichtlich in den nächsten 5 Jahren grundlegend ändern. Daher ist wachsende Aufmerksamkeit auch für die negativen Seiten von QML gefragt.

Funding. Open Access funding enabled and organized by Projekt DEAL.

Literatur

[1] Morgan, S. "Global cybercrime damages predicted to reach $6 trillion annually by 2021," Cybersecurity ventures. https://cybersecurityventures.com/annual-cybercrime-report-2020/ (Erstellt: 26. Okt. 2020). Zugegriffen: 28. Feb. 2022.

[2] Alani, M. M. (2019). *Applications of machine learning in cryptography: a survey*. ICCSP '19. : Association for Computing Machinery.

[3] Liu, W., Huang, P., Peng, J., Fan, J., & Zeng, G. (2018). Integrating machine learning to achieve an automatic parameter prediction for practical continuous-variable quantum key distribution. *Physical Review A*, *97*. https://doi.org/10.1103/PhysRevA.97.022316

[4] Lau, H.-K, & Lo, H.-K. (2011). Insecurity of position-based quantum-cryptography protocols against entanglement attacks. *Physical Review A*. https://doi.org/10.1103/PhysRevA.83.012322

[5] Beigi, R. K. S. (2011). Simplified instantaneous non-local quantum. *New Journal of Physics*. https://doi.org/10.1088/1367-2630/13/9/093036

[6] Bykovsky, A. Y. (2021). Multiple-valued logic and neural network in the position-based cryptography scheme. *Journal of Russian Laser Research*. https://doi.org/10.1007/s10946-021-10000-7

[7] Alani, M. M. (2012). *Neuro-cryptanalysis of des and triple-des*. International Conference on Neural Information Processing.

[8] Shor, P. (1994). *Algorithms for quantum computation: discrete logarithms and factoring*. Proc. Annual Symp. on Foundations of Computer Science.

[9] Keplinger, K. (2018). Is quantum computing becoming relevant to cyber-security? *Network Security, 16–19*. https://doi.org/10.1016/S1353-4858(18)30090-4

[10] Galer, S. (2020). Quantencomputer bedrohen die Unbezwingbarkeit von Blockchains. https://news.sap.com/germany/2020/06/quantencomputer-blockchain/ Zugegriffen: 28. Feb. 2022.

[11] Krempl, S. (2021). „Abhörsicher: Erste quantengesicherte Videokonferenz zwischen Bundesbehörden," Heise online. https://www.heise.de/news/Abhoersicher-Erste-quantengesicherte-Videokonferenz-zwischen-Bundesbehoerden-6159925.html Zugegriffen: 28. Feb. 2022

[12] Yiyu Mao, Wenti Huang, Hai Zhong, Yijun Wang, Hao Qin, Ying Guo, Duan Huang (2020). Detecting quantum attacks: a machine learning based defense strategy for practical continuous-variable quantum key distribution. *New Journal of Physics*. https://iopscience.iop.org/article/10.1088/1367-2630/aba8d4

[13] Al-Mohammed, , et al. (2020). *Machine learning techniques for detecting attackers during quantum key distribution in IoT networks with application to railway scenarios*. IEEE Access.

[14] N. I. o. S. a. Technology (2020). NIST's post-quantum cryptography program enters 'selection round'. https://www.nist.gov/news-events/news/2020/07/nists-post-quantum-cryptography-program-enters-selection-round Zugegriffen: 28. Feb. 2022.

[15] Gast, R. „Überlegenheit der Quanten: Fünf Fragen zu Googles Quantencomputer," Spektrum.de. https://www.spektrum.de/news/ueberlegenheit-der-quanten-fuenf-fragen-zu-googles-quanten-computer/1681398 (Erstellt: 23. Okt. 2019). Zugegriffen: 28. Feb. 2022.

[16] Arute, F. (2019). Quantum supremacy using a programmable superconducting processor. *Nature*. https://doi.org/10.1038/s41586-019-1666-5

[17] Lee, Y., Joo, J., & Lee, S. (2019). Hybrid quantum linear equation algorithm and its experimental test on IBM Quantum Experience. *Scientific Reports*. https://doi.org/10.1038/s41598-019-41324-9

[18] Wittek, P. (2014). Supervised learning and support vector machines. In *Quantum machine learning*. : Academic Press.

[19] Rebentrost, P., Mohseni, M., & Lloyd, S. (2014). Quantum support vector machine for big data classification. *Physical Review Letters*. https://doi.org/10.1103/PhysRevLett.113.130503

[20] Senekane, M., Mafu, M., & Taele, B. (2017). *Privacy-preserving quantum machine learning using differential privacy*. 2017 IEEE AFRICON.

[21] GPT-3 A robot wrote this entire article. Are you scared yet, human. https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3 (Erstellt: 8. Sept. 2020). Zugegriffen: 28. Feb. 2022.

[22] Bosnjak, Z., Grljevic, O., & Bosnjak, S. *CRISP-DM as a framework for discovering knowledge in small and medium sized enterprises*. SACI 2009. Published 28 May 2009.

[23] Wang, X., Li, J., Kuang, X., Tan, Y.-A., & Li, J. (2019). The security of machine learning in an adversarial setting: a survey. *Journal of Parallel and Distributed Computing*. https://doi.org/10.1016/j.jpdc.2019.03.003

[24] Huang, L., Joseph, A. D., Nelson, B., Rubinstein, B. I. P., & Tygar, J. D. (2011). *Adversarial machine learning*. Proceedings of 4th ACM Workshop on Artificial Intelligence and Security.

[25] Bursztein, E. (2018). Attacks against machine learning—an overview. https://elie.net/blog/ai/attacks-against-machine-learning-an-overview Zugegriffen: 28. Feb. 2022.

[26] Tramèr, F., Zhang, F., Juels, A., Reiter, M. K., & Ristenpart, T. (2016). *Stealing machine learning models*. USENIX Security Symposium.

[27] Biggio, B., & Roli, F. (2018). Wild patterns: ten years after the rise of adversarial machine learning. *Pattern Recognition*. https://doi.org/10.1016/j.patcog.2018.07.023

[28] Lu, S., Duan, L.-M., & Deng, D.-L. (2020). Quantum adversarial machine learning. *Phys. Rev. Research*. https://doi.org/10.1103/PhysRevResearch.2.033212

[29] Liu, N., & Wittek, P. (2020). Vulnerability of quantum classification to adversarial perturbations. *Phys. Rev. A*, *101*. https://doi.org/10.1103/PhysRevA.101.062331

[30] Liao, H., Convy, I., Huggins, W. J., & Whaley, K. B. (2021). Robust in practice: Adversarial attacks on quantum machine learning. *Physical Review A*. https://doi.org/10.1103/PhysRevA.103.042427

[31] Wiebe, N., & Kumar, R. S. S. (2018). *Hardening quantum machine learning against adversaries*

[32] Bang, J., Lee, S.-W., & Jeong, H. (2015). Protocol for secure quantum machine learning at a distant place. *Quantum Information Processing*. https://doi.org/10.1007/s11128-015-1089-7

[33] IBM (2021). Real quantum computers. https://quantum-computing.ibm.com/ Zugegriffen: 28. Feb. 2022.

[34] Barreno, M., Nelson, B., Sears, R., Joseph, A. D., & Tygar, J. D. (2006). *Can machine learning be secure?* ASIACCS'06.