深層学習を用いた人工知能暗号システムの設計と解析

研究代表者　　櫻　井　幸 一　　九州大学大学院　システム情報科学研究院　情報学部門

1 はじめに

暗号通信を行う際、第三者の盗聴防御を防ぐ手段として、人工知能自らが自律的に、暗号アルゴリズムを設計し、現実に利用できる次世代暗号通信の理論を開拓する。人工知能が構築した暗号通信システムの強度も、攻撃者が人工知能であると想定しての解析と評価を行う敵対的機械学習モデルを適用する。本研究では、先駆的な Google 研究者の敵対的機械学習理論が、人工知能時代の安全な暗号通信の実現に期待できることが動機となり、Google 暗号理論を掘り下げ、未解決である公開鍵暗号系への拡張問題にも取り組む。

暗号は、機密保護通信の基礎要素として古代から利用され、公開鍵暗号の発明により、公衆回線上で、安全なインターネットショッピングができる時代である。これまでの暗号は、人間が設計し、その安全性は、数理学を用いた暗号解析理論によって、議論されてきた。またニューラルネットを利用した暗号系も、多くの研究者が提案してきたが、その多くは時間を待たずに解読され、その限界を示している。

サイバーセキュリティでは、人口知能ロボットがインターネットや IoT 機器を攻撃する時代になって、その防御も人口知能で対応するか、という課題が議論されつつある。

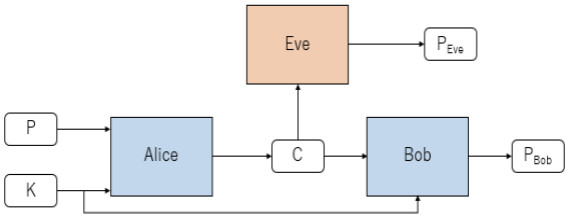
この研究は、人間が設計した暗号アルゴリズムを使って、人口知能ロボット同士が秘匿通信を行うのではなく、自律的にロボット自らが、人工知能により生成した暗号を利用しての秘密通信を行い、盗聴防御を防ぐという次世代秘密通信を研究するものである。

2 背景と目的

背景：Google 研究所の Abadi と Andersen は、暗号通信方式の設計と解析に、敵対的機会学習の理論を適用した新たなモデルを提案した[Abadi and Andersen. Learning to protect communications with adversarial neural cryptography (2016)]。彼らは、共通鍵暗号方式において、利用するニューラルネットは原始的ではあるが、人間に検知されることなく、ロボット間で自律的に安全な通信が可能であることを示し、人工知能による次世代暗号通信の画期的な研究と評価できる。原論文の付録では、公開鍵暗号系への拡張も議論されているが成功には至らず、現在まで未解決の挑戦課題である。

それまでのニューラル暗号では、人間がニューラルネットを使って設計し、暗号通信に利用する。それを盗聴者である人間が、数理解析を使って、暗号解読を行う。その解読の困難性で、設計された暗号の安全性強度を評価する、という流れであった。しかし、Abadi らは、暗号の設計も解読も全て、人工知能だけが行うという斬新な設計解析モデルを提案した。

本研究では、深層学習、特に敵対的機械学習理論を、暗号アルゴリズムの設計と解析にどこまで適用できるか、その応用範囲を広げ、また限界を明らかにすることを目標とした。具体的には、（1）Google の２者間共通鍵暗号通信を、グループ通信へ拡張すること、(2) さらには公開鍵暗号まで拡張が可能かどうかを研究すること、　（3）また、暗号通信だけでなく、電子透かしや難読化など情報秘匿型通信への展開も掘り下げること、の３つの事項である。



公益財団法人電気通信普及財団
研究調査助成報告書　第38号　2023年度

3 本研究調査に関する内外の研究の動向

Google 以前：ニューラルネットを利用した暗号の設計は、古くから研究されてきた。しかし、そのほとんどが解読されている[Klimov, Mityagin, and Shamir “Analysis of Neural Cryptograpy” Asiacrypt 2002]。Google 暗号理論(2006)以前は、設計と解析とが分離していた：設計と構成にニューラルネットを応用するが、その解析評価は人間が頭脳で行うもの。いくつかも暗号モジュールから、暗号アルゴリズムを自動合成し、その強度解析は人間が行うもの、であった。

Google 暗号に刺激された研究：国内では、会津大の蘇が、中国の研究グループと Google 暗号の解析を発展されている[Zhou, Chen, Yeh, Zhang, and James, “Security analysis and new models on ..”]。また、代表者の櫻井グループが、進化ネットワークへ Google 暗号理論を適用した研究だけである。

海外でも、Google 暗号を直接に掘り下げた研究は[Coutinho et.al. ” Learning perfectly secure cryptography to protect communications with adversarial neural cryptography. Sensors (2018)]と、画像の透かしへ適用した研究[Yedroudj et al. “Steganography using a 3 player game” (2019)]の２件だけであった(本研究計画申請時の 2021 年 11 月)。

Google 暗号の研究展開は、まだ一部の研究者に限られている。しかし、そのアイデアの元となった敵対的機械学習の理論は、2014 年に Goodfellow らが提案して以来、深層学習の分野でもっと活発な研究分野の 1 つである。

また、Google 暗号の提案者の一人である Abadi は、30 年前から暗号理論を形式的手法により研究してきたが、最近は、暗号に加えて、データプライバシー保護への深層学習の応用も発表している[Deep Learning with Differential Privacy, ACM-CCS 2016]。

Google 研 Abadi-Anderson の論文：

> 概要の抄訳：本研究では、ニューラルネットワークが、別のニューラル ネットワークからの情報を保護するために、秘密鍵の利用法を学習できるかどうかを明らかにする。具体的には、マルチエージェント システムにおける機密性の確保に焦点を置き、それらの特性を敵対者の観点から特定するものである。このため、システムはアリスとボブという名の 2 つのニューラル ネットワークで構成され、イブという名の 第三のニューラル ネットワークが、アリスとボブの間の通信を盗聴して学習しようとすることを抑止制限することを目的とする。我々のモデルでは、これらのニューラル ネットワークに特定の暗号アルゴリズムを規定はしない。代わりに、敵対的にエンドツーエンドで訓練した結果のアルゴリズムが対象となる。本研究では、ニューラル ネットワークが、暗号化と復号化の両形式を実行する方法、さらに機密性を達成するためにこれらの操作を選択的に適用する方法を学習できることを実証する。

2016 年に Arxiv で公開された原論文は、深層学習分野のトップ国際会議である The International Conference on Learning Representations (ICLR)2017 に投稿されていたことが確認できる。この分野では、査読過程でのレビューコメントや判定意見まで、公開されていることに注意する。しかし、ICLR2017 では、採択に至っておらず、Abadi-Anderson の原論文は、Arxiv で公開された版を引用できる現状である。しかし、この画期的な論文が公開できたおかげで、我々が本研究を始めるきっかけとなっている。また、原論文の内容は、決して完全なものとは言い難く、逆にこのことが、更なる研究課題を提供してくれている。

Google Scholar によると、Abadi-Anderson の原論文は、250 件程度の引用がある (2023 年 6 月現在)。M. Abadi ら Google 研が 2015 年に発表した TensorFlow 論文の引用 26,900 件以上に比すると、250 件の引用件数は小さい数値ではある。しかし、暗号分野の引用研究としては、十分に引用さて、継続研究を触発していると評価できる。ちなみに、暗号研究の大家である Adi Shamir が 2019 年に発表した敵対的攻撃の理論解明に関する研究[A simple explanation for the existence of adversarial examples with small hamming distance]の引用が 80 件程度である。Shamir は、2021 年に、さらに敵対的攻撃に関する数理解析論文[The dimpled manifold model of adversarial examples in machine learning]も発表しているが、この引用は

30 件程度であり、Shamir 自身の前者の研究論文も引用されていない。いずれにせよ、正式な査読のある会議や論文誌に掲載されずとも、Arxiv で公表されれば、その内容精査と引用しての発展研究は、読者である研究者自身の裁量とも言える時代になってきたとも言える。

人工知能暗号の国際学会動向:2021 年より欧州国際暗号会議 Eurocrypt 併設ワークショップの 1 つとして、Artificial Intelligence and Cryptography(AICrypt)が始まっている。以下は、ワークショップ HP からの抄訳である。

近年、人工知能（AI）とセキュリティの相互作用がより顕著かつ重要になってきた。セキュリティをより効率的に向上させる必要があり、当然のことである。 AI 応用が着実に増加しているセキュリティ分野の 1 つは暗号化である。 AI 技術が実装攻撃、PUF への攻撃、ハードウェアトロイ木馬の検出などをどのように改善できるかは、既に明らかになってきた。

暗号化における AI の役割に加え、 AI のための暗号化が新たな重要なトピックであると考えられる。 AI システムに対する攻撃の数が増加しているため、考えられる研究の 1 つの方向性は、そのような脅威を軽減するためにどの暗号技術を使用できるかを調査することも必要であろう。この AICrypt ワークショップでは、暗号化と AI のさまざまな側面に取り組む学界と産業界の研究者を集め、その経験を共有し、連携を強化する方法を議論することを目的とする。我々は、多様な暗号化アプリケーションと AI 保護メカニズムの間での技術の転用可能性を調査研究することに特に興味を持っている。

主要課題：

- 深層学習による暗号解読（例: ニューラル識別子）
- 暗号解読のための AI モデルの説明可能性と解釈可能性 -
- サイドチャネル分析のための深層学習技術
- 暗号原理とプロトコルの AI 支援設計
- 暗号プロトコルに対する AI 主導の攻撃
- AI システムのセキュリティとプライバシーのために暗号論的対策

また、さらに 2022 年より国際暗号化会議 Crypto の併設ワークショップとして

The Glowing Hot Topics in Cryptography (GHTC, 暗号における白熱する話題)

が始まっている。2022 年は、暗号の機敏性(Cryptographic Agility)が主題であったが、2023 年この夏予定の主題は 「暗号と AI の出会い」と宣言している：

暗号技術における最先端の話題を取り上げることを趣旨とする本ワークショップ GHTC の第 2 回目のテーマは「Crypto meets AI」である。このテーマは、暗号と人工知能の接点を探り、セキュリティ、プライバシー、その他デジタル通信の重要な側面を強化するために、この 2 つの分野がいかに協力できるかを探索することを目的とする

しかし、本研究で取り上げている主題である Abadi-Anderson 発の人工知能暗号に関する研究発表は、上記の AICrypt でも、未だ見当たらない現状にある。

新研究会の発足と活性化：人工知能学会/安全性とセキュリティ研究会

暗号に限らず、信頼性やサイバーセキュリティも含む人工知能周辺の研究の交流と活性化のために、人工知能学会に(第二種)研究会を、2022 年 4 月に下名が主導し発足させた。

2022 年 11 月 23 日 オンラインと慶応大でのハイブリッド開催の集会を開催できた:

開催主旨: 近年、AI 技術を利用した多くの製品やサービスが世の中に浸透してきており、AI の意思決定が人々の生命や多くの産業に影響を与えるものになっている。AI による自律的な意思決定から 人間が徐々に排除されていく中で、設計原理として AI のセキュリティを考慮する必要性が高まっている。本研究会では、AI のセーフティとセキュリティに関する誤動作、攻撃、防御、追跡、 分析を含む新しいアイデアを広く模索し、研究を深めることを目的とする。

2023年は、発足二年目にして、“人工知能学会誌 Vol.38 No.2 (2023年3月発行)に「AI セキュリティの研究動向」の特集を組むことにも成功した。

4 Abdai-Anderson 原論文手法の課題

Abadi と Andersen による研究は、多くの前向きなフィードバックと継続研究が見られる一方で、対処すべき複数の欠陥があった。

(F1)最も重要な欠陥は、生成される暗号文の安全性である。生成された暗号文は、$\chi 2$ 攻撃のような複数の統計的攻撃に弱いことが示されている。また、暗号文は米国政府 NIST の統計テストにも不合格であった。

(F2)この提案のもう一つの問題は、一般的なニューラルネットワークモデルの非凸性に関連している。 学習中にモデルが最小化しようとする目的の損失関数は凸型ではないことは知られている。ある関数が一意的な大域的最小点を持つ場合、その関数は凸であると言われる。しかし、深層学習をはじめ Abadi-Anderson のモデルでは、損失関数は非常に複雑な非凸型である。この非凸性の問題は、2 つの別々のニューラルネットワークが同じデータで学習しても、それぞれが異なる局所極小点に収束する可能性が高いため、同じ暗号化／復号化技術を学習することが保証されないことを意味する。このため、2 者以上の間での通信は、対処すべき難しい問題となる。

(F3) Abadi と Andersen の原論文では、ニューラルネットワークは暗号化と復号化に対称鍵しか使えないと制限である。実際には、Abadi と Andersen の原論文の付録に記載されているが、ニューラルネットワークに非対称鍵の使用を学習させようとした試みは成功には至らず、受信側は送信されたメッセージを復号することができなかった報告がある。

(F4) 最後の欠点は、これまでの欠点ほど深刻ではないが、ニューラルネットワークの訓練に要する時間である。GAN の訓練は時間と電力を消費することが知られており、電力に制限のあるデバイスがモデルを使用するのは難しい。

上記の 4 つの課題(F1)～(F4)に関しては、既存研究で指摘された事項もあり、また、一部の解決案も提案刺されていた。しかし、本研究では、特に(F2)の課題は、ニューラルネット応用における基本問題として認識したことが重要であったと評価している。

5 本研究の主要結果

Ishak が代表者指導の博士課程で始めた研究をきっかけに、本助成研究プロジェクトを通じて、先に掲げた課題(F1)～(F4)に対して、我々は、以下のような解を与えた。

(S1) 弱い暗号文の問題や、非凸性の問題による複数当事者間の通信の困難性に対処するため、我々は、複数の当事者（ニューラルネットワーク）が完全に安全な対称暗号（ワンタイムパッド暗号）を学習し、ランダムに初期化された複数のニューラルネットワークが同期して、攻撃者から通信を保護するために使用される同じ暗号技術を学習することを可能にするニューラルネットワークモデルを提案した。これらの攻撃者(イブ)は、選択平文攻撃のような既知のさまざまな暗号解読攻撃を適用するが、通信当事者（アリスとボブ）は攻撃が成功するたびにペナルティを受け、これらの攻撃に対して堅固なより強力な暗号文を生成するように働く。我々の完全安全な暗号化ニューラルネットワークモデルは、生成される暗号文の安全性を向上させるための様々な提案に基づいている。また、当事者のサブグループが非公開で通信する方法や、新しい訓練セッションを初期化することなくグループ通信を実行する方法も与えている。さらに、このモデルを用いて、

任意の一般的なアクセス構造を実現する秘密分散方式を得た。

(S2) Jamie Hayes と George Danezis[16]によって提案されたステガノグラフィ・ニューラルネットワークモデルも、先に述べた非凸性の問題を有している。Jamie Hayes と George Danezis はこの点を強調し、非凸性の問題により、異なるマシン上で学習する 2 つのニューラルネットワークが同じステガノグラフィアルゴリズムを学習できることは保証されないと彼らの原論文で言及している。我々は、この問題に対処するため、2 つ以上のニューラルネットワークが同じステガノグラフィアルゴリズムを学習できる方法を与えた。このアプローチは、我々の多者間敵対的暗号化アルゴリズムから着想を得ており、2 つ以上のニューラルネットワークが同じステガノグラフィアルゴリズムを学習し、３者間の場合に焦点を当てて、画像内に隠されたメッセージを正しく抽出する方法を与えた。

(S3) ニューラルネットが非対称鍵を使ってデータを暗号化／復号化できないという問題に対処するため、本研究では、非対称情報を使って暗号化を学習するという、この種のものとしては初めてのニューラルネットモデルを提案した。非対称暗号化により、ニューラルネットワークは一対の鍵を用いて直接通信することができる。最初の鍵（公開鍵）は誰でも知っており、送信するメッセージや文書を暗号化するために使用する。2 番目のキー（秘密キー）は受信者だけが知っており、メッセージや文書の復号化を可能にする。既存技術では、あらかじめ別の暗号プロトコルを使って共有された鍵で、ニューラルネットワークがデータを暗号化・復号化する必要があった。また、我々は、提案手法が、複数の暗号解読攻撃に対して安全であることも検証した。

(S4) IoT デバイスのような性能に制限のあるデバイスに最適でない可能性のある長い学習時間の問題に対処するために、そのようなデバイスに適した代替の軽量ニューラルネットワークモデルを提案した。このモデルは、I. Kanter、W. Kinzel、E. Kanter[40]によって提案された Tree Parity Machines（木型パリティ機械、TPMs）鍵交換プロトコルに基づく安全な暗号化方式を学習するものでる。このモデルは、NIST の統計テストに合格するに十分なランダム性を持つ暗号文を生成するように学習する。このモデルは Abadi と Andersen のモデルよりも軽量であり、学習も大幅に高速である。Abadi と Andersen のモデルとは対照的に、TPM 鍵交換プロトコルは、隠れ層が 1 つしかないシンプルなニューラルネットワークモデルを使用しているため、学習が高速である。このモデルはヘッブ学習技法を用いて学習し、GANs セットアップにおける敵対的学習には依存しない。肝心の安全性に関しては、TPM 鍵交換が既に過去に何度も攻撃を受け、安全でなかったことが指摘されていたため、鍵の推測を困難にするような TPM 鍵交換プロトコルへの最新の貢献に基づいき鍵交換プロセスを実行した。[注意：TPMs モデルは、Abadi と Andersen のような GAN ではなく、Hebbian 学習に基づいている。]

成果発表論文

数年前より、人工知能暗号の設計と解析を主題に研究に取り組んできた博士学生 Ishak と共同研究者を中心に、今回は主要成果を 2 つの学術論文で発表する成果を得た。

[R1] Learning asymmetric encryption using adversarial neural networks
(敵対的ニューラネットワークを用いた非対称暗号の学習)
by Ishak Meraouche, Sabyasachi Dutta, Haowen Tan, Kouichi Sakurai
Engineering Applications of Artificial Intelligence
Pub Date: 2023 , DOI:10.1016/j.engappai.2023.106220

概要/抄訳：この研究では、送信者であるアリスと受信者であるボブが、複数の攻撃盗聴者であるイブらから通信を保護するために、1 組の公開鍵と秘密鍵の対を利用を学習とする、マルチエージェント型敵対的ニューラルネットワークモデルを提案する。この分野の既存研究では、学習過程を開始する前に、アリスとボ

ブの間で対称情報を共有する必要があった。我々の知る限り、今回の我々の研究は、非対称情報を持つアリスとボブが通信を保護するための訓練が行える最初の試みであると評価している。我々のモデルでの初期設定においては、送信者アリス、受信者ボブ、盗聴者イヴ、そして 2 つの（公開鍵生成器と秘密鍵生成器と呼ぶ）ニューラルネットワークの 5 つのエージェントから構成される。ボブからの（秘密の）ランダムノイズに基づいて、公開鍵を使い、秘密メッセージを復号化することを、イヴの解読を阻止しながら、アリスが公開鍵でメッセージを暗号化し、ボブが秘密鍵でメッセージを復号化することを可能にする公開鍵と秘密鍵のペアを生成する。我々は、ニューラルネットワークが通信を確立し、イブの盗聴からも保護できることを示す。最後に、漏洩攻撃、選択平文攻撃をモデル化し、暗号文の識別可能性を試験評価するために、イブよりも強い敵対者をも考慮した。最後の 3 つの実験により、(非対称情報を持つ) ニューラルネットワークがより強固な安全性と、秘密鍵からの漏洩を含む漏洩攻撃に対する耐性をも有し、秘密通信を保護できることも示した。

[R2] "Learning Multi-Party Adversarial Encryption and Its Application to Secret Sharing,"
(多者間敵対的暗号学習とその秘密分散への応用)
by I. Meraouche, S. Dutta, S. K. Mohanty, I. Agudo and K. Sakurai,
in *IEEE Access*, vol. 10, pp. 121329-121339, 2022,
DOI: 10.1109/ACCESS.2022.3223430.

概要/抄訳：ニューラル・ネットワークに基づく暗号は、敵対的生成ネットワーク Generative Adversarial Networks, GAN) を利用して暗号を学習できるニューラル・ネットワークを構築する Google の敵対的暗号の導入以来、改良と応用に関する研究が続いている。Google 暗号は、そう安全性強度が高くないことも、当初から指摘されていた。しかし、多くの後続研究のおかげで、ニューラルネットワークに使い捨て機能(ワンタイムパッド, One Time Pad) を学習させ、完全に安全な暗号文を生成できることも示されている。我々の知る限り、既存研究では、2 者から、せいぜい 3 者間の通信しか考慮していなかった。これに対して、今回の研究では、敵対的な初期設定における複数のニューラルネットワークが、異なる攻撃者に通信を盗聴されている状況下で、いかに遠隔で同期し、完全に安全な通信を確立できるかを示すことに成功した。その 1 つの応用として、完全に安全な複数者間通信に基づく秘密分散方式を構築する方法も与える。その結果、4 つのニューラルネットワークが同期し、平衡に達するまでにおよそ 45,000 回程度の学習ステップを必要とすることも実験的に明らかにした。平衡に達したとき、すべてのニューラルネットワーク同士は互いに暗号通信することができ、攻撃者はネットワーク間で交換された暗号文を解読することは困難である。

5 本研究の体制

研究グループの構成：代表が櫻井、分担は蘇（会津大)、分担補助博士学生 Ishak の国内グループに加え、2019 年 8 月まで NICT 海外プログラムにより招聘していたインド人ポスドク Saby. Dutta(当時・カルガリー大、現在はインド TSG センター)が海外外部アドバイサーとして協力した。

準備：既存研究調査：分担補助 Ishak とアドバイサーDutta とは、それまでの 1 年半年をかけて、ニューラル暗号の調査を行ってきた。特に、Google の敵対的ニューラル暗号理論(2016)の発表以前には、暗号設計にニューラルネットを活用する多数の論文が発表されている。この調査をまとめた論文は IEEE Access に採択されオープンアクセス可能である。

Ishak Meraouche, Sabyasachi Dutta, Haowen Tan, Kouichi Sakurai:
Neural Networks-Based Cryptography: A Survey.
IEEE Access 9: 124727-124740 (2021)

Iskak の博士コース研究課題：国費留学生である Ishak(アルジェリア)は、本研究を、学位論文の主要テーマとして推敲した。分担の蘇は、九州大学から正式に博士外部アドバイサーとして、共同研究に加えて、学位取得に向けた研究アドバイスもお願いしていた。

さらに良い状況の変化としては、人工知能の暗号やサイバーセキュリティへの応用を含むテーマで、一昨年申請していた JSPS-DST 日印国際共同研究が採択され、2022 年 6 月から2年間の予定で進行中である。この共同研究の一環もあって、インド側研究者 Mohanty 博士(IIIT-dmj)も、論文投稿にあたって助言討論を持って、共著者となった。2023 年 12 月には、来日しての対面での共同研究討論も予定している。

また、2022 年 9 月からは、スペインはマラガ大の Roman 博士が、母国政府の支援を受けて、一年間下名の研究室に滞在している。彼にも、IEEE-access 論文への再投稿にあたって、助言と支援をお願いできた。さらに、同僚で数学系出身の Isaac 博士も紹介してもらい、実質的な共同研究・共著者となった。マラガ大では、2023 年から、サイバーセキュリティー学部コースを発足させ、そこに人工知能も交えた卒業プログラムを計画しているということ。

Ishak は、一連の研究成果により、博士号の学位を取得予定である（2023 年 9 月)。

6 さいごに：今後の課題

9 月に学位取得予定の Ishak の博士論文は、九州大学学位論文書誌データベースよりオープンアクセス可能となる。さらに、Ishak が博士論文をまとめる過程の討議で、この分野の未解決課題も洗い出すことができた。

主要課題(F3)であった公開鍵暗号の実現に関しては、成果(S3)で 1 つの解を与えたが、使い捨て(one-time pad)情報を用いるものであって、Diffie-Hellmann 流の公開鍵暗号系実現という本来の課題は以前と未解決問題である[文献[30]の著者である Zhu Yuetong の九州大学修士論文(2019)での考察]。

特に、アリスとボブの送受信者が、初期設定で共有すべき情報がどの程度必要であるかの限界解明である。この問題は暗号理論においても、ほとんど手付かずの研究課題と見ている。こうした考察は、研究代表である下名が、今年度暗号系シンポジウムで発表することを考えている。

注：2017 年 8 月、Facebook の人工知能研究機関 Facebook AI Research が行ったある実験が話題になったことがある。2 つの AI(Chatbot)アリスとボブで会話実験を行なった。すると、観測している人間が理解できない言語で会話となり、実験は強制終了されたというニュースである。これに関しては、世界のメディアが「ついに AI が意思を持ち人間を脅かすのでは」と報じていた。AI の暴走は、米マイクロソフト社の Toy の事件も記憶にあるが、こちらは人間に仕業と言われている。

ChatGPT の登場とその脅威が議論され始めた今日、人工知能同士の自律的言語による会話（それは人間が理解できない言語）の能力の可能性と限界を解明する時期が来たとも考える。

【参考文献】

[1] R. Spillman, M. Janssen, B. Nelson, and M. Kepner, “Use of a genetic algorithm in the cryptanalysis of simple substitution ciphers,” Cryptologia, vol. 17, no. 1, pp. 31–44, Jan. 1993.

[2] N. D. Truong, J. Y. Haw, S. M. Assad, P. K. Lam, and O. Kavehei, “Machine learning cryptanalysis of a quantum random number generator,” IEEE Trans. Inf. Forensics Security, vol. 14, no. 2, pp. 403–414, Feb. 2019.

[3] J. So, “Deep learning-based cryptanalysis of lightweight block ciphers,” Secur. Commun. Netw., vol. 2020, Jul. 2020, Art. no. 3701067.

[4]F.Wang, R.Ni, J.Wang, Z.Zhu, X.Chen, and Y.Hu, “Novel fully convolutional network for cryptanalysis of cryptosystem by equal modulus decomposition,” Laser Phys. Lett., vol. 17, no. 9, Jul. 2020, Art. no. 095201.
[5] T.-W. Yue and S. Chiang, “A neural network approach for visual cryptography,” in Proc. Int. Joint Conf. Neural Netw., vol. 5, Jul. 2000, pp. 494–499.
[6] W. Yu and J. Cao, “Cryptography based on delayed chaotic neural networks,” Phys. Lett. A, vol. 356, nos. 4–5, pp. 333–338, Aug. 2006.
[7] A. Klimov, A. Mityagin, and A. Shamir, “Analysis of neural cryptography,” in Advances in Cryptology (Lecture Notes in Computer Science), vol. 2501, Y. Zheng, Ed. Queenstown, New Zealand: Springer, Dec. 2002,
pp. 288–298, doi: 10.1007/3-540-36178-2_18.
[8] M.Abadi and D.G.Andersen, “Learning to protect communications with adversarial neural cryptography,” CoRR, vol. abs/1610.06918, pp. 1–15, Oct. 2016.
[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in Proc. Adv. Neural Inf. Process. Syst., vol. 27, 2014.
[10] Y. Gao, R. Singh, and B. Raj, “Voice impersonation using generative adversarial networks,” 2018, arXiv:1802.06840.
[11] Y.Kataoka, T.Matsubara, and K.Uehara,“Image generation using generative adversarial networks and attention mechanism,” in Proc. IEEE/ACIS 15th Int. Conf. Comput. Inf. Sci. (ICIS), Jun. 2016, pp. 1–6.
[12] A. Siarohin, S. Lathuiliere, E. Sangineto, and N. Sebe, “Appearance and pose-conditioned human image generation using deformable GANs,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 4, pp. 1156–1171, Apr. 2021.
[13] L. Zhou, J. Chen, Y. Zhang, C. Su, and M. A. James, “Security analysis and new models on the the intelligent symmetric key encryption,” Computer Security, vol. 80, pp. 14–24, Jan. 2019.
[14] M. Coutinho, R. de Oliveira Albuquerque, F. Borges, L. G. Villalba, and T.-H. Kim, “Learning perfectly secure cryptography to protect communications with adversarial neural cryptography,” Sensors, vol. 18, no. 5, p. 1306, Apr. 2018.
[15] M. Yedroudj, F. Comby, and M. Chaumont, “Steganography using a 3-player game,” J. Vis. Commun. Image Represent., vol. 72, Oct. 2020, Art. no. 102910.
[16] J. Hayes and G. Danezis, “Generating steganographic images via adversarial training,” in Proc. Adv. Neural Inf. Process. Syst., vol. 30, 2017, pp. 1954–1963.
[17] I. Meraouche, S. Dutta, and K. Sakurai, “3-Party adversarial cryptography,” in Advances in Internet, Data and Web Technologies, L. Barolli, Y. Okada, and F. Amato, Eds. Cham, Switzerland: Springer, 2020, pp. 247–258.VOLUME 9, 2021 124739
[18] D. V. Vargas and J. Murata, “Spectrum-diverse neuro evolution with unified neural models,” 2019, arXiv:1902.06703.
[19] Z. Li, X. Yang, K. Shen, R. Zhu, and J. Jiang, “Information encryption communication system based on the adversarial networks foundation,” Neurocomputing, vol. 415, pp. 347–357, Nov. 2020.
[20] É. S. Dorokhin, W. Fuertes, and E. Lascano, “On the development of an optimal structure of tree parity machine for the establishment of a cryptographic key,” Secur. Commun. Netw., vol. 2019, pp. 1–10, Mar. 2019.
[21] T. Dash, S. N. Dambekodi, P. N. Reddy, and A. Abraham, “Adversarial neural networks for playing hide-and-search board game Scotland yard,” Neural Comput. Appl., vol. 32, no. 8, pp. 3149–3164, Apr. 2020.

[22] A.Ruttor, W.Kinzel, and I.Kanter, “Dynamics of neural cryptography,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 75, no. 5, May 2007, Art. no. 056104.
[23] O. M. Reyes and K.-H. Zimmermann, “Permutation parity machines for neural cryptography,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 81, no. 6, Jun. 2010, Art. no. 066117.
[24] L. F. Seoane and A. Ruttor, “Successful attack on permutation- parity-machine-based neural cryptography,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 85, no. 2, Feb. 2012, Art. no. 025101.
[25] J.JinandK.Kim,“3DCUBE algorithm for the key generation method: Applying deep neural network learning-based,” IEEE Access, vol. 8, pp. 33689–33702, 2020.
[26] N. Prabakaran and P. Vivekanandan,“A new security on neural cryptography with queries,” Int. J. Adv. Netw. Appl., vol. 2, pp. 437–444, 2008.
[27] S. Kiyomoto, H. Ota, and T. Tanaka, “On-the-fly automatic generation of security protocols,” in Proc. ICEIS, vol. 2, Jan. 2008, pp. 97–104.
[28] H.Ota, S.Kiyomoto, and Y.Miyake,“Automatic security verification for 3-party authentication and key exchange protocols,” in Proc. 5th Int. Conf. Netw. Syst. Secur., Sep. 2011, pp. 254–258.
[29] J.Purswani, R.Rajagopal, R.Khandelwal ,and A. Singh,“Chaos theoryon generative adversarial networks for encryption and decryption of data,” in Proc. Adv. Bioinf., Multimedia, Electron. Circuits Signals. Springer, 2020, pp. 251–260.
[30] Y. Zhu, D. V. Vargas, and K. Sakurai, “Neural cryptography based on the topology evolving neural networks,” in Proc. 6th Int. Symp. Comput. Netw. Workshops (CANDARW), Nov. 2018, pp. 472–478.
[31] W. Zheng, K. Wang, and F. Wang, “GAN-based key secret-sharing scheme in blockchain,” IEEE Trans. Cybern., vol. 51, no. 1, pp. 393–404, Jan. 2020.
[32] A. Narayanan and V. Shmatikov, “Robust de-anonymization of large sparse datasets,” in Proc. IEEE Symp. Secur. Privacy, May 2008, pp. 111–125.
[33] E.S.Finn, X.Shen, D.Scheinost, M.D.Rosenberg, J.Huang, M.M.Chun, X. Papademetris, and R. T. Constable, “Functional connectome finger- printing: Identifying individuals using patterns of brain connectivity,” Nature Neurosci., vol. 18, no. 11, pp. 1664–1671, Nov. 2015.
[34] C. Huang, P. Kairouz, X. Chen, L. Sankar, and R. Rajagopal, “Context- aware generative adversarial privacy,” 2017, arXiv:1710.09549. [Online]. Available: http://arxiv.org/abs/1710.09549
[35] J. Shi, S. Chen, Y. Lu, Y. Feng, R. Shi, Y. Yang, and J. Li, “An approach to cryptography based on continuous-variable quantum neural network,” Sci. Rep., vol. 10, no. 1, Dec. 2020, Art. no. 2107.
[36] D. P. Kingma and J. Ba, “Adam: A method for stochastic opti- mization,” 2014, arXiv:1412.6980.
[37] Y. Ke, M. Zhang, J. Liu, and T. Su, “Generative steganography with Kerckhoffs’ principle based on generative adversarial networks,” CoRR, vol. abs/1711.04916, pp. 1–8, Nov. 2017.
[38] V.Dunjko,J.M.Taylor,andH.J.Briegel,“Quantum-enhancedmachine learning,” Phys. Rev. Lett., vol. 117, no. 13, Sep. 2016, Art. no. 130501, doi: 10.1103/PhysRevLett.117.130501.
[39] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, “Supervised learning with quantum- enhanced feature spaces,” Nature, vol. 567, no. 7747, pp. 209–212, 2019.
[40] I. Kanter, W. Kinzel, and E. Kanter, “Secure exchange of information by synchronization of neural networks,” Europhys. Lett., vol. 57, no. 1,pp. 141–147, Jan. 2002.

研究代表者のグループ発表論文

[G1] 3-Party Adversarial Cryptography,
Ishak Meraouche, Sabysachi Dutta, Kouichi Sakurai,
EIDWT-2020; LNDECT Volume 47 pp. 621-626 (January 2020)

[G2] 3-Party Adversarial Steganography
Ishak Meraouche, Sabysachi Dutta, Kouichi Sakurai
WISA-2020; LNSC Volume 12583 pp. 89-100 (December 2020)

[G3] Neural Networks-Based Cryptography: A Survey
Ishak Meraouche, Sabysachi Dutta, Haowen Tan; Kouichi Sakurai
IEEE Access; Volume 9 pp. 124727 - 124740 (September 2021

[G4] Key Exchange Using Tree Parity Machines: A Survey
Ishak Meraouche, Kouichi Sakurai
ICAIAA-2021; Springer AIS 2524-7565 pp. 363 - 372(February 2022)

[G5] Tree Parity Machine based Symmetric Encryption: a Hybrid Approach
Ishak Meraouche, Sabysachi Dutta; Kouichi Sakurai
ICMC-2022; Springer Mathematics and Computing pp. 1—11 (February 2023)

[G6] Learning Multi-Party Adversarial Encryption and Its Application to Secret Sharing
Ishak Meraouche, Sabyasachi Dutta, Sraban Kumar Mohant, Isaac Agudo,
and Kouichi Sakurai
IEEE Access; Volume: 10 pp. 121329 - 121339 (November 2022)

[G7] Learning asymmetric encryption using adversarial neural networks
Ishak Meraouche, Sabysachi Dutta, Haowen Tan, Kouichi Sakurai
Engineering Applications of Artificial Intelligence, Volume 123, Part B, 106220,
(Online available on May 2023)

〈発　表　資　料〉

題　　名	掲載誌・学会名等	発表年月
Learning asymmetric encryption using adversarial neural networks	Engineering Applications of Artificial Intelligence, Volume 123, Part B, Elsevier	2023-Aug. DOI: 10.1016/j.engappai.2023.106220
Learning Multi-Party Adversarial Encryption and Its Application to Secret Sharing	*IEEE Access*, vol. 10, pp. 121329-121339,	2022, doi: 10.1109/ACCESS.2022.3223430.
Science of Cyber Security - SciSec 2022 Workshops : AI-CryptoSec, TA-BC-NFT, and MathSci-Qsafe 2022, Matsue, Japan, August 10–12, 2022, Revised Selected Papers Editors: Chunhua Su, Kouichi Sakurai	Communications in Computer and Information Science, Springer Nature Singapore	2023/01/02 DOI https://doi.org/10.1007/978-981-19-7769-5