

未来社会創造事業 探索加速型
「超スマート社会の実現」領域
年次報告書(探索研究)

令和元年度 研究開発年次報告書

令和元年度採択研究開発代表者

[研究開発代表者名：藤野 毅]

[立命館大学 理工学部・教授]

[研究開発課題名：エッジ AI のハードウェアセキュリティに関する研究]

実施期間：令和元年 11 月 1 日～令和 2 年 3 月 31 日

§1. 研究開発実施体制

(1)「立命館大学」グループ

① 研究開発代表者: 藤野 毅 (立命館大学理工学部, 教授)

② 研究項目

(A) 攻撃対象となるエッジ AI エンジンの FPGA 実装/MPU 実装

(1) シストリックアレイを用いた AI エンジンの FPGA 実装

(3) サイドチャネル攻撃対策技術の実装

(B) エッジ AI ハードウェアに対する網羅的な攻撃手法と対策手法の研究

(1) サイドチャネル情報を用いた AI モデルのリバースエンジニアリング攻撃と対策

(3) サイドチャネル情報を用いたモデル抽出攻撃と対策

(4) イメージセンサインターフェースへの Adversarial Examples 攻撃と対策

(5) エッジ AI デバイスへのフォルト攻撃と対策

(C) AI 自身のセキュリティ全般に関する研究キャッチアップ

(2)「三菱電機」グループ

① 主たる共同研究者: 中井 綱人 (三菱電機 情報技術研究所, 研究員)

② 研究項目

(A) 攻撃対象となるエッジ AI エンジンの FPGA 実装/MPU 実装

(2) AI エンジンの MPU を用いた実装

(3) サイドチャネル攻撃対策技術の実装

(B) エッジ AI ハードウェアに対する網羅的な攻撃手法と対策手法の研究

(2) サイドチャネル情報を用いた Adversarial Examples 攻撃と対策

§2. 研究開発実施の概要

目標とする、セキュリティを強化したエッジ AI 処理ハードウェアの実現にあたって、研究項目 (A) に示した攻撃対象となるエッジ AI エンジンの開発が必要であり、FPGA 実装 (A-1) 並びに/MPU 実装 (A-2) を行った。FPGA 上の実装にあたっては、Deep Neural network (DNN) のアクセラレータとしてよく利用されるシストリックアレイをサイドチャネル攻撃評価用の ZUIHO ボード上に実装した。MPU 上の実装にあたっては、ARM プロセッサ搭載 Mbed ボード上に AI フレームワーク (TensorFlow) を利用して C++ 言語で実装した環境を構築した。

上記で構築したエッジ AI エンジンを用いて、研究項目 (B) にあたる、サイドチャネル情報を用いた AI モデルのリバースエンジニアリング攻撃 (B-1) とサイドチャネル情報を用いた Adversarial Examples (AE) 攻撃 (B-2) と対策 (B-3) を行った。前者では、基礎検討用の 3×3 演算器を搭載したシストリックアレイ回路が動作している時の消費電力波形をオシロスコープで観測し、相関電力解析を用いて解析した結果、ターゲットである DNN モデルパラメータ 9 個全てを導出することができ

た. 後者では, 入力データに応じた AI の処理時間差をサイドチャネル情報として利用することで, AE が作成可能であるというタイミング攻撃を発見した. さらに, このタイミング攻撃が活性化関数の処理時間差に起因することを見出し, 活性化関数の処理時間を一定化する手法を提案, 実装したところ, タイミング攻撃を行っても, ランダムノイズと同等のノイズしか発生できず, AE 生成に対する対策となっていることを確認した.

研究項目(C)に示した AI 自身のセキュリティ全般に関する研究キャッチアップに関しては, 「出力確率の量子化による対策がされた DNN に対する Black-Box Adversarial Attack 手法」の提案と, 「深層学習におけるバックドア攻撃に対する蒸留を用いたバックドア除去手法」の提案を行い, それぞれ国際会議で発表を行ったところ, 高い評価を受け両論文とも Student Paper Award を受賞した.