

TEE を用いたセキュアな機械学習環境へ向けた一検討

平間海渡 中林舜葵 光澤敦
秋田県立大学 システム科学技術学部 情報工学科

関する現状や課題を明らかにする。

1 はじめに

機械学習の利用方法としてクラウド上で学習・推論を行い、出力を端末で得るという方法と、手元の端末に学習モデルを保持させ推論・学習を行うという2つの方法がある。クラウド上で利用する場合の利点として、端末の処理能力に依存しないことや通信部分を信頼できるものにすればデータが漏れにくいという点があるが、クラウドと端末間での通信時間がオーバーヘッドとして存在することで即応性のあるタスクには用いにくいという欠点がある。手元の端末に学習モデルを保持させる場合は、ネットワークの接続性が存在しない環境においても学習・推論を行うことができるという利点があるが、端末からの学習モデルの盗難・改ざんが成される危険があるという欠点がある。

学習モデルを盗難や改ざんから保護する方法として、信頼できる実行環境である Trusted Execution Environment(TEE) を用いるという方法が存在する。しかし、現在の TEE では利用できるメモリ量に制限があり大規模な機械学習・推論を行うには不十分であることから、学習モデルを分割する手法を用いた機械学習・推論が提案されている。

本研究では手元の端末に学習モデルを保持した場合の学習モデル保護について着眼し、Arm 社の TEE である Arm TrustZone(TZ)[1] とオープンソースの SecureOS である Open Portable TEE(OP-TEE)[2]、機械学習・推論フレームワークである Darknet[3] を用いて、エッジデバイスでも実行可能な小さな学習モデルにおける推論の実行時間を測定し、TEE を用いた機械学習環境に

2 関連研究

Mo ら [3] は深層学習フレームワークである Darknet を拡張し、TZ で動作可能な DarknetTZ を実装した。DarknetTZ では、1 つの機械学習モデルを学習層ごとに分割することが可能であり、その分割単位ごとに機械学習・推論を行う実行環境を通常の実行環境(REE) と TEE から指定することができる。学習層を分割することによって、プライバシー情報の特徴が現れやすい後半層のみを TEE で実行し、それ以外の層をより計算資源が潤沢な REE で実行することが可能となる。

Bayerl ら [4] は OP-TEE の改変を行うことによって、端末のマイクやカメラ等のデータ収集を行う周辺機器に対しても TEE を用いた保護が可能な手法を提案している。これにより、機械学習モデルのみでなく機械学習・推論時に取得するデータに対しても保護がなされるため、よりセキュアな機械学習・推論を行うことが可能となる。

3 実験

本研究では、事前に作成された機械学習モデルに対して TEE を用いた推論を行った場合にかかる学習モデル読み込み時間、推論実行時間の2つを測定した。

3.1 実験環境

TEE として TZ を用い、SecureOS として OP-TEE を、機械学習フレームワークとして DarknetTZ を用いた。測定用に作成した機械学習モデルの構造は LeNet であり、MNIST データセットを用いて作成した。

3.2 実験方法

事前準備として画素数が 14×14 , 28×28 , 56×56 の MNIST データセットを作成し、それぞれの画素数のデータセットに対して 0~10 層の計 11 層からなる LeNet 構造の機械学習モデルを作成した。この際、データセットの画素数と等しくなるようにそれぞれのモデルへの入

Secure execution of a machine learning model using TEE
Hirama Kaito, Nakabayashi Mitsuki, and Mitsuzawa Atsushi

Akita Prefectural University Faculty of Systems Science
and Technology Department of Information and Computer Science

力値を調整した。作成した機械学習モデルを DarkneTZ を用いて REE 実行部と TEE 実行部に分割し、推論実行時間を測定した。機械学習モデルの分割の内訳は、1)0~10 層の全層を TEE, 2)0 層を REE, 1~10 層を TEE, 3)0~4 層を REE, 5~10 層を TEE, 4)0~8 層を REE, 9~10 層を TEE, 5) 全層を REE, の5つとした。

学習モデル読み込み時間の測定 学習モデルの読み込み処理の範囲は、学習モデルの構造と学習パラメータが入力されてから、暗号化された学習パラメータの復号やメモリの確保が終了し変数として出力されるまでとした。

推論実行時間の測定 読み込まれた学習モデルを用いて推論を行うが、推論時間として測定する処理の範囲は分割されたモデルへの対象画像の入力から、その推論結果が返されるまでとした。推論処理後に行われる推論結果のソートなどの実行時間は測定範囲外とした。

3.3 実験結果

学習モデルの読み込み時間に関する結果を図1に、推論実行時間に関する結果を図2に示す。読み込み時間、推論実行時間ともに全層の処理を TEE で実行した場合が最も実行時間が大きくなり、全層を REE で実行した場合が最も実行時間が小さくなった。

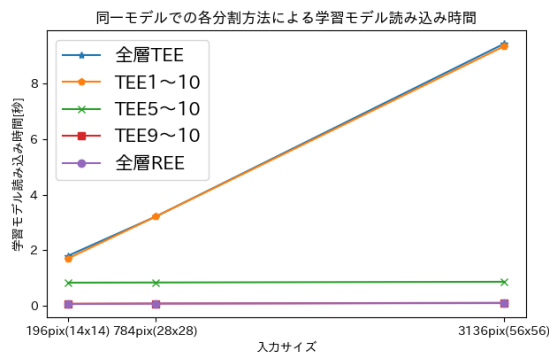


図1 同一モデルで TEE 実行部範囲による学習モデル読み込み時間の変化

4 結論

本研究では、TEE の内部に展開可能な小さな学習モデルを用いて、学習モデルの読み込み時間と推論実行時間を測定した。図1から分かるように今回用いた機械学習環境では、機械学習モデルの読み込みに要する時間が主な実行時間オーバーヘッドとなり主な課題になり得

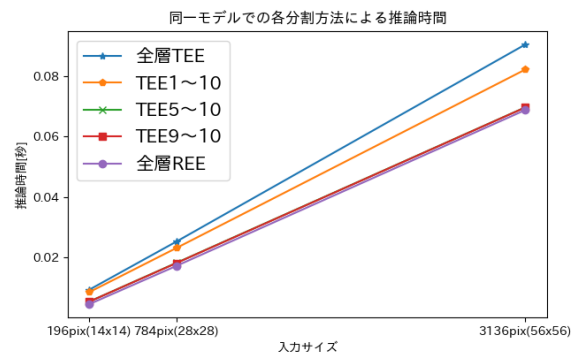


図2 同一モデルでの TEE 実行部範囲による推論時間の変化

ることがわかる。しかし、機械学習モデルを TEE に展開した状態で運用することによりモデル読み込みのオーバーヘッドを考慮せずに運用することが期待できる。また図2から、用いる学習モデルのサイズが小さい場合にはネットワークを介した推論タスクよりも小さな応答時間で推論結果を得ることが可能である。よって現在のエッジデバイスにおける TEE の機械学習環境についても、推論タスクを行う対象の選定を行うことにより、実用することができるといえる。

参考文献

- [1] ARM, : ARM Security Technology Building a Secure System using TrustZone Technology.
- [2] Open Portable Trusted Execution Environment - OP-TEE, <https://www.op-tee.org/> Accessed: 2022/12/28 15:34:19.
- [3] Mo, F., Shamsabadi, A. S., Katevas, K., Demetriou, S., Leontiadis, I., Cavallaro, A. and Haddadi, H.: DarkneTZ, in *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, ACM (2020).
- [4] Bayerl, S. P., Frassetto, T., Jauernig, P., Riedhammer, K., Sadeghi, A.-R., Schneider, T., Stapf, E. and Weinert, C.: Offline Model Guard: Secure and Private ML on Mobile Devices, in *2020 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, IEEE (2020).