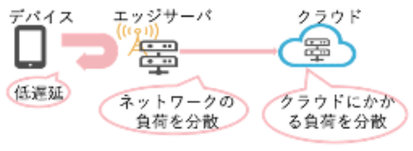
小口研究室　研究紹介　(2021年度)
(お茶の水女子大学理学部情報科学科)

リッチデバイスを用いたプライバシ保護に優れた分散機械学習モデルにおける顔画像認識
（研究担当:高野 紗輝）

研究背景

◆ スマートフォンやIoTデバイスの普及
◆ 機械学習の活用機会が増加 → エッジデバイスで収集した大量のデータに対してプライバシを保護しながら複雑な処理を行うことが求められる

◆ デバイスに近いネットワークエッジにエッジサーバを配置して処理を行うエッジコンピューティングが注目されている
◆ しかし、性能の低いエッジデバイス側はあくまでデータを収集し、そのデータをエッジサーバに転送するという役割



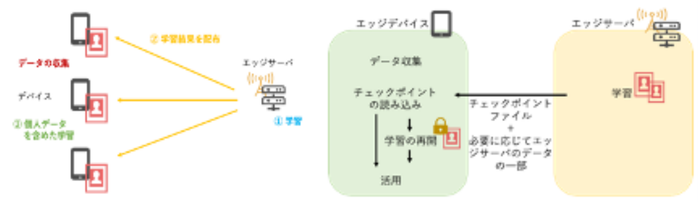
研究目的

◆ 研究課題
1. プライバシの保護
2. エッジサーバに渡すことができないエッジデバイス上のデータを含めた機械学習

◆ 研究目的
複雑な処理も行うことのできるリッチデバイスの登場
⇩
リッチデバイスに適した分散機械学習モデルの構築

提案モデル

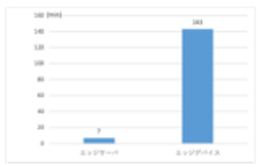
◆ エッジコンピューティングにおいて従来エッジサーバ上で行っていたタスクの一部をエッジデバイスにオフロードする



- エッジサーバで一般的なデータを用いて学習した結果をエッジデバイスに配布(重みを保存したチェックポイントファイルを送信)
- エッジデバイスの持つ個人データとエジサーバから受け取った一般的なデータで学習を再開

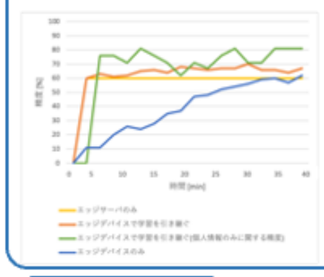
実験

◆ エッジデバイスとエッジサーバの性能比較
エッジデバイスとして、GPU を搭載した小型 AI コンピュータボードであるJetson Nanoを使用



ある程度低速だがサーバと同等の精度まで学習が進み、エッジデバイス内のみでも十分学習可能

◆ 提案モデルの実装
顔画像データセットであるlfwから32人(30枚ずつ)用意し、2割をテストデータ、残りを訓練データとしてぼかし等により9倍にして使用



<エッジサーバの全てのデータを受け取る>
・エッジデバイス上における個人データを含む学習によって全体の精度および個人情報に対する精度が上がる
・エッジサーバの助けを借りることが有効

◆ エッジサーバから受け取るデータの割合を変化

エッジサーバから受け取る一般的なデータの割合を0～10割まで変化させた際の全体の精度

個人情報のみに対するの精度

エッジデバイス上での学習には
1. 個人データの他に一般的なデータが必要
2. エッジサーバのデータのうち1割を与えるだけでかなり良い精度となる
3. 3割与えると上限近くまで学習可能

今後の課題

◆ 個人情報に対応したより良いモデルの検討
◆ プライバシを保護した上でエッジデバイスの情報をエッジサーバやクラウドにフィードバックするモデルの検討

完全準同型暗号データマイニングアプリケーションにおける高性能SSD活用の考察
（研究担当:廣江 彩乃）

研究背景

- **暗号化アプリケーションの重要性**
  医療情報など、秘匿データを扱う際には暗号化アプリケーションが必要不可欠
- **完全準同型暗号**
  暗号化したまま加算・乗算の演算が可能であるという、大変便利な手法であるが、暗号演算処理が重く、実用化には程遠い
- **クラウドへの委託**
  特に暗号化アプリケーションを実行する際、演算処理が多い上に、扱うデータが多いことから、クラウドを用いることが想定される
  クラウドの使用の際には、用いるコンピュータリソースによってコストに大きな差があるため、コンピュータリソースのコストパフォーマンスが大変重要
- **高性能SSD**
  低コスト、低レイテンシかつ大容量な不揮発性のSSDの開発が進む

図：一般的なアプリケーションと完全準同型暗号を用いた暗号化アプリケーション

実験概要

完全準同型暗号を用いた暗号化アプリケーション実行に高性能SSDを用い、性能評価を行う。
演算処理の多い暗号化アプリケーション実行時にメインメモリが不足する際、HDDなどのストレージにswapが行われることが一般的である。
しかし、メインメモリに比べるとストレージへのアクセス速度は遅く、これが大きなボトルネックとなっている。
そこで、高性能SSDをswap領域として用いることで、実行の様子がどのように変わるかを調査する。

実験

【調査内容】
アプリケーションのコンピュータリソースへの負荷比較
•実行時間
•swap発生状況

【実験環境】
実験で用いたサーバと記憶装置について、下の二つの表に示す

表：実験に用いたサーバについて

Server	
CPU	Intel® Xeon® Processor E5-2643 v3 6 Cores × 2 Sockets
DRAM	DDR4, 512GB, 2133MT/s
HDD	HGST, SATA, 8TB

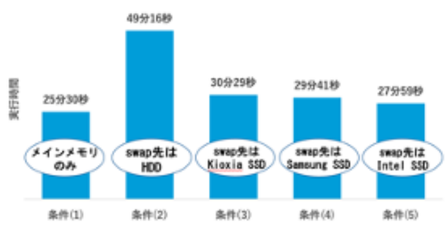
【対象アプリケーション】
卒業生・山本(2018)による購買データ秘匿データマイニングアプリケーション
購買データから特定のアイテムセットを含む購買データのマイニングを行う、クライアントサーバ構成アプリケーション

表：実験に用いた記憶装置

	KIOXIA EXCERIA PLUS SSD	SAMSUNG SSD 980 PRO	Intel Optane SSD 800P Series
容量	1TB	500GB	118GB
搭載メモリ	NAND フラッシュ	NAND フラッシュ	3D XPoint

実験結果

【実験結果】
・アプリケーションの実行時間は、特に低レイテンシである3DXPointを用いたIntelのSSDを用いた条件が最短
・swap速度はSSDの読み書き性能よりも遅い
→SSDのレイテンシが実行時間に大きく影響することがわかった



Android端末における無線環境と有線環境を考慮したTCP実装による通信制御
（研究担当:松野 瑛南）

研究背景

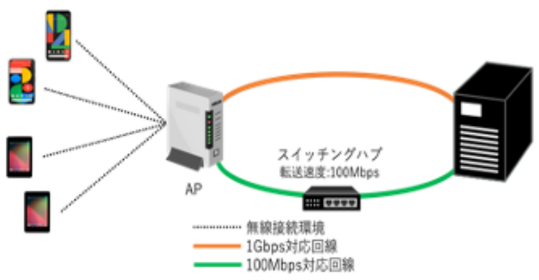
- **スマートフォン市場の拡大**
  - 端末が高性能になり、ヘテロな環境が発生
- **輻輳発生原因の増加**
  従来：有線の通信速度が圧倒的に速い
  ↓
  近年：有線環境より無線環境の方が速くなり、有線側がボトルネックになる

通信環境がより複雑に

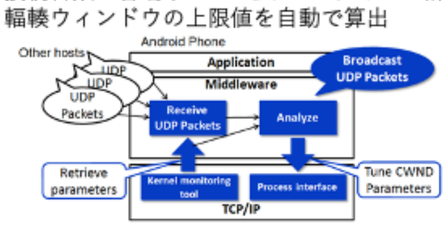
→ 先行研究とは異なる環境
合計通信速度/公平性から評価
↓
Fairness Indexの改良が必要
提案、評価を行う

実験概要

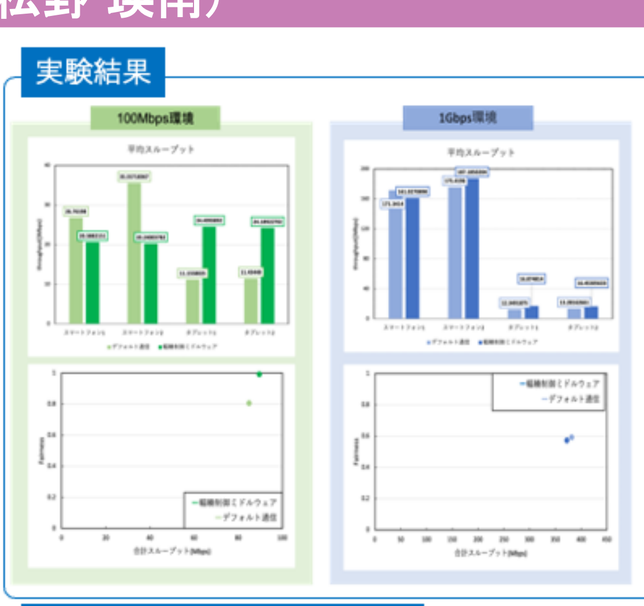
iperfを用いて複数のAndroid端末とサーバ間の50秒間通信を行う
- 端末4台の同時通信
- デフォルト通信と輻輳制御ミドルウェアの比較



- 輻輳制御ミドルウェア
・接続台数と各端末のRTTをリアルタイムに計測
・輻輳ウィンドウの上限値を自動で算出



実験結果



- 100Mbps環境
  - 公平性、合計通信速度が向上
- 1Gbps環境
  - 期待通りに動作しない

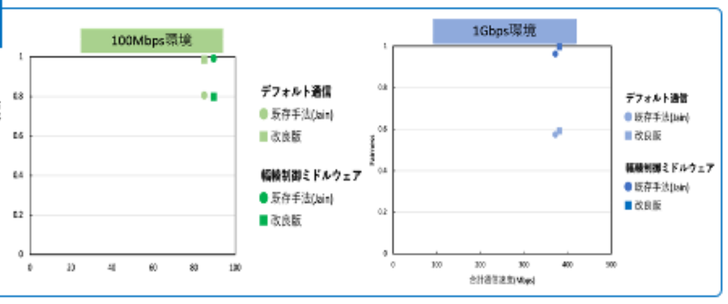
Fairness Indexの問題点

・取得したスループットの値を使用
・通信性能に大きな差がある場合正しく評価できない

改良版Fairness Indexの提案

通信性能に**重み**をつけた公平性の指標

$$F_j(X_n) = \frac{\left(\sum_{i=1}^{n} \frac{x_i}{Max_i}\right)^2}{n\sum_{i=1}^{n} \frac{x_i}{Max_i}^2} (1 \le k \le 1)$$



http://ogl.is.ocha.ac.jp/
oguchi@is.ocha.ac.jp