小口研究室　研究紹介　(2017年度)
(お茶の水女子大学理学部情報科学科)

完全準同型暗号を用いたゲノム秘匿検索の高速化手法
（研究担当：山田　優輝）

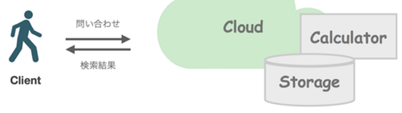
研究背景

◆ ゲノムデータ委託システム

大型のストレージと計算機を所有するクラウドにデータと計算を委託
→ 膨大なゲノムデータを用いた統計処理を行うことができる
特定の文字列がゲノム配列に含まれているかどうか判定する問い合わせを行う

◆ プライバシ保護・計算量

個人のゲノムデータは住所などとは異なり変更することができない
→ 暗号化によるプライバシー保護が必要
膨大なデータを暗号化して処理するため計算量が課題となる



提案手法

◆ 完全準同型暗号 FHE を利用

暗号文同士での加算と乗算が成立する完全準同型暗号 FHE を利用
メリット: クラウド上で復号せずに演算を行うことが出来るため
サーバ・クライアント双方のデータを秘匿することが出来る
デメリット: FHE は計算量が大きいためサーバ側での計算時間が長くなってしまう
演算を行いすぎると復号することが出来なくなる

◆ マスタ・ワーカ型分散

サーバ側にデータベースの分割による分散処理を導入
→ 一度により多くのデータとクエリのマッチングを
調べることが出来るようになる
演算回数の上限を除去するためブートストラップ法も導入

提案システム概要

実験と考察

◆ 平均実行時間測定実験

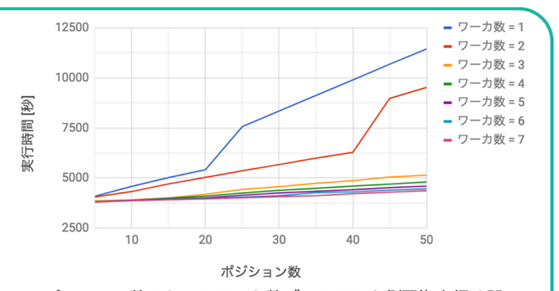
- ポジション: 検索開始位置（複数指定可）
- ポジション数とワーカ数を変えて秘匿検索演算の実行時間を測定
- 高速化率 ＝ 逐次実行時間 / 並列実行時間 を算出

◆ 実行時間

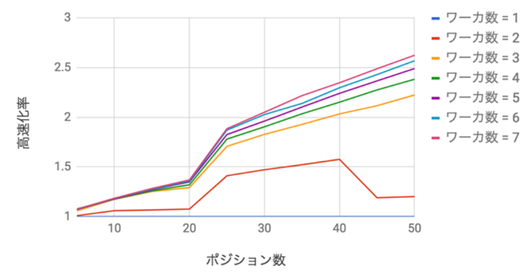
- ワーカ数の増加に伴い実行時間が減少する
- 実行時間は段々と横ばいになる
- ポジション数を増やすと実行時間が増加する

◆ 高速化率

- ワーカ数・ポジション数の増加に伴い高速化率も高くなる
- ポジション数が多いほど分散化効果が大きい
  - ダミーのポジションを含む検索が効果的



ポジション数におけるワーカ数ごとのマスタ側平均実行時間



ポジション数におけるワーカ数ごとの高速化率

今後の課題

◆ 検索対象のサンプル数を増やした実験を行う
◆ 更なる高速化に取り組む
  ◆ 実用に向け頻繁に行われる操作に特化した高速化を検討
◆ プライバシレベルのコントロールを検討

無線通信端末のパケットの深層学習を用いた解析
（研究担当：山本　葵）

研究背景

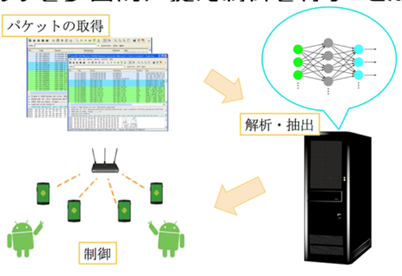
- スマートフォン、タブレット端末などのワイヤレスデバイスの普及
- トラフィックの増加

→無線LANの負荷、帯域の取り合いによる輻輳の発生

研究目的

- 無線LANのトラフィックの深層学習による解析
- 解析によるトラフィックの予測
  →輻輳の極めて早期な検出、予兆の発見

トラフィックを多面的に捉え制御を行うことが目標



パケットの取得

AirPcap

- 米国Riverbed社の製品
- Wireshark 統合型のワイヤレストラフィックパケットキャプチャデバイス
- 制御、管理、データの各フレームを含む、低レベルの IEEE802.11 a/b/g/n ワイヤレストラフィックをキャプチャ

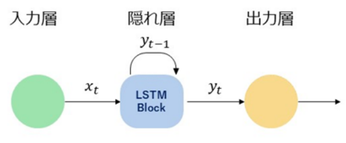
深層学習

Chainer

深層学習用フレームワーク
Preferred Networks 社が開発

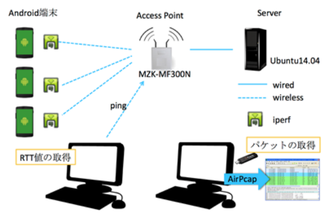
LSTM(Long Short-Term Memory)

時系列データに対するモデル
隠れそうのユニットをLSTM blockと呼ばれるメモリと3つのゲートを持つブロックに置き換えることで実現



実験環境

アクセスポイントに接続した複数のAndroid端末から
iperfによってデータをサーバに送信、データの取得

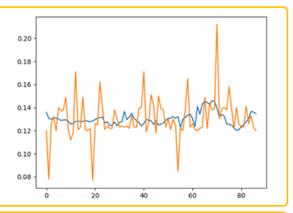


実験1

t ~ t+9の入力データからt+10のRTT値を予測
データ(87秒)は同じで入力データの次元のみ変更

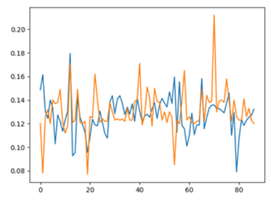
入力データ
- 1秒間の平均データ量

正解データ
- RTT値



入力データ
- 時間
- パケット数
- 平均データ量
- 送信機器の台数
- 受信機器の台数

正解データ
- RTT値



今後の課題

- まとめ
  - 入力データの次元の変更により精度が向上した
  - 長時間データではあまり精度は向上しない
- 今後の課題
  - 多次元データ(1パケットずつの入力)による実験

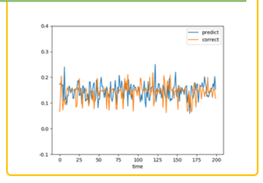
実験2

t ~ t+9の入力データからt+10のRTT値を予測
入力データの次元は同じで長時間の学習による精度評価
汎化能力の検証のため学習用データとテスト用データを用意

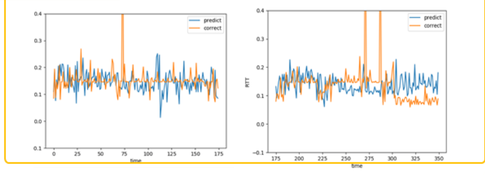
入力データ
- パケット数
- 平均データ量
- 送信機器の台数
- 受信機器の台数

正解データ
- RTT値

学習データ(609秒)(グラフは 前半200秒)



テストデータ(353秒)



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