5ZC-02

完全準同型暗号を用いた秘匿データマイニングと高性能SSDによる性能向上に関する考察

廣江　彩乃†　圓戸　辰郎††　小口　正人†

† お茶の水女子大学
†† キオクシア株式会社

1 はじめに

近年，個人情報を含むビッグデータを扱う処理をクラウドコンピュータに委託する取り組みが増えている．企業などの様々な組織は，顧客から個人情報や技術情報などの機密情報を保持しており，解析技術の発展に伴って，これらのデータを利用することでさらに有用なデータを得ることができる．一方で，クラウド上でこれらのデータから解析結果を得る際には，処理依頼データの送信から演算途中，そして演算結果を送信するまでの過程において，データ漏洩リスクがある．例えばゲノムデータなどの生体情報や医薬品の開発における技術情報は利用価値が高く，盗聴対象となりやすいためデータを秘匿したまま処理を行うことができる完全準同型暗号 [1] が重要である．
しかしこの手法には，コンピュータリソースへの負荷が大きすぎるという問題がある．アルゴリズムの高速化は進められているものの，依然としてサーバ側の秘匿計算の量が多く，メインメモリの使用量が多くなるためである．そこで本研究では，近年開発が進む高速な SSD を活用することを考える．複数の SSD やメインメモリを秘匿検索に用いて，その際のコンピュータリソースへの負荷や実行時間といった実行状況の比較・評価を行うことで，効率よく秘匿検索手法を実行するためのメモリの利用方法を検証する．また，完全準同型暗号方式を用いると暗号化データが元の数万倍のデータ量になるため，大概メインメモリが不足するが，メモリが高価であることを考慮すると，実行時間の課題に加えてコストの面にも課題がある．
そこで実行効率とコストの課題を解決するため，近年高速化に向けて研究開発が進む，高性能で比較的安価な SSD の有効利用を検討する．

2 先行研究

本研究では先行研究で実装された完全準同型暗号による Apriori アルゴリズムによるデータマイニングプログラムを用いる．Apriori アルゴリズムはデータベースのアイテム間における相関ルール抽出の効率的な手法として，1993 年に Agrawal ら [2] によって提案され，アソシエーション分析などの頻出データマイニングで用いられる有名なアルゴリズムである．購買履歴を対象とする場合，トランザクションごとに商品の種類である各アイテムを購入したか否かをバイナリ表現で記録したデータベースから，各アイテムの部分集合であるアイテムセットの頻度計算を行う．頻度は全トランザクション数に対して対象のアイテムセットの全アイテムを購入したトランザクションの数の割合であり，これをサポート値と表現する．Apriori アルゴリズムは，探索するアイテムセットの長さを徐々に増加させながら，幅優先的に繰り返しサポート値計算と枝刈りを行うことで頻出なアイテムセットを導出する．アプリケーションの仕組みは図 1 に示す．
完全準同型暗号を用いた安全委託頻出パターンマイニング手法としては，Liu ら (2015) [3] によって P3CC(Privacy Preserving Protocol for Counting Cadidates) が提案された．Liu らが提案した安全委託システムは，トランザクションごとに購入したかどうかを 0 か 1 でバイナリ表現された購買データについて Apriori 計算を行うサーバ・クライアント型のシステムとして設計されている．この従来手法に対して高橋ら (2015) [4] や今林ら (2016) [5] によって P3CC 手法の計算時間の大幅な改善が行われた．一方で，完全準同型暗号は暗号文同士の演算を行うため依然として計算量は多く，実用化に向けてはまだ実行時間の長さの課題が残る．

3 実　験

3.1 実験概要

本研究では上で述べた完全準同型暗号を用いた秘匿データマ

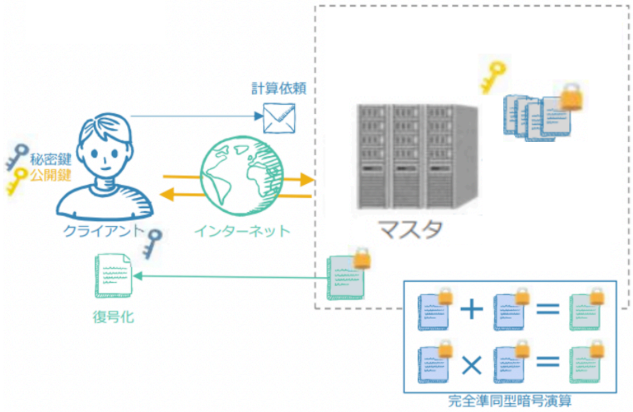

図 1　データマイニングアプリケーションの仕組み

Consideration of privacy-preserving data mining with FHE by using high performance SSDs
†Ayano Hiroe　††Tatsuro Endo　†Masato Oguchi
†Ochanomizu University
††Kioxia Corporation

イニングアプリケーションを用いて実行時間を計測し，実行環境の差による分析を行う．
プログラム実行に必要なメインメモリの量が容量を超える際には，ストレージに領域を確保する必要がある．この場合において，メインメモリに比べると圧倒的に遅いストレージへのアクセスが大きなデータを扱う暗号化アプリケーションの実行時間にどの程度影響するか調べるため，swap 処理に着目する．本実験ではクラウド環境を想定し，使用可能なメインメモリを制限して，メインメモリの外の swap 領域へのアクセスを発生させる．そして，その swap 先メモリに高性能 SSD を指定して，アクセス速度の差を検証していく．用いたサーバのスペックは表 1 に示した．

3.2 高性能 SSD の比較

swap デバイスに高性能な SSD を用いた場合の性能評価を行っていく．比較対象とする高性能 SSD は 3 種類である．それらの性能をまとめたのが表 2 である．表 2 の 3 種類の SSD を swap 先として指定して実行する他に，実行に十分なメモリを割り当てることで swap 処理を発生させない条件 (1) と，メモリが不足する状況を模倣して HDD に swap 領域を作成させる条件 (2) を加えて，表 3 に示す 5 種類のケースで計測していく．

3.3 実験結果及び分析

秘匿データマイニングアプリケーションを DRAM 上のみで実行を終わらせた場合と，使用可能メモリを制限して swap の状況を作った場合で負荷を取得し，比較した．表 3 に示した条件ごとに実行時間を計測し，その結果を図 2 に示す．

表 1 サ ー バ

CPU	Intel®Xeon®Processor 6 Cores × 2 Sockets
DRAM	DDR4 512GB 2133MT/s
HDD	HGST SATA 2TB
PCIe Gen	3.0

表 2 比較対象 SSD

	Kioxia EXCERIA PLUS SSD [6]	Samsung 980 PRO [7]	Intel OPTANE SSD 800P [8]
capacity	1TB	500GB	118GB
memory type	NAND Flash	NAND Flash	3D XPoint

表 3 実 行 条 件

	メイン メモリ	HDD	Kioxia SSD [6]	Samsung SSD [7]	Intel SSD [8]
条件 (1)	○	-	-	-	-
条件 (2)	○ (※)	○	-	-	-
条件 (3)	○ (※)	-	○	-	-
条件 (4)	○ (※)	-	-	○	-
条件 (5)	○ (※)	-	-	-	○

(※) cgroup によって使用可能メモリを 0.2GB に制限

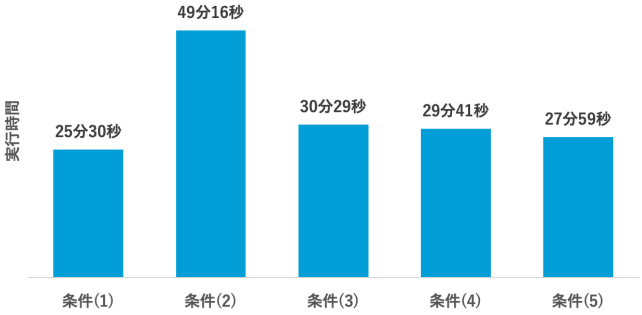

図 2 実 行 時 間

実行時間を比較すると，メインメモリの制限がない条件 (1) が実行時間が最短で，swap 領域として HDD を用いた条件 (2) は最長であった．そして，swap 領域として SSD を用いた条件 (3) から条件 (5) を比較すると，読み書きの遅延であるレイテンシが短い特徴を持つ 3DX Point というメモリを搭載する SSD を使用した条件 (5) が条件 (3),(4) と比べて顕著に実行時間が短い結果となった．SSD を用いた条件の結果を見ると，スワップ処理ではストレージのレイテンシ差が性能差となり，HDD を用いた場合と比べて DRAM とのハードウェアによる大きな性能差は見られないことが分かった．

4 まとめと今後の課題

今回行った実験結果より，高性能 SSD を用いることで，暗号化アプリケーション実行時に発生した swap 処理を高速化できることがわかった．今後，性能差がある SSD を用いるとパフォーマンスに違いがあるか，どのような性能を持つデバイスが暗号化アプリケーションに適しているかを調べていく．また，完全準同型暗号を用いた暗号化アプリケーションにおいて，どのような特徴や処理内容を持つものに対してレイテンシの低さなどの高性能 SSD 性能が有効であるのかを分析し，実用化に向けて研究を進めていく．

謝　　辞

本研究の一部は，キオクシア株式会社と JST CREST JPMJCR1503 の支援を受けて実施したものである．

文　　献

[1] Craig Gentry, et al., Fully homomorphic encryption using ideal lattices. In STOC, Vol. 9, pp. 169–178, 2009
[2] Rakesh Agrawal, Tomasz Imielinski, and Arun Swami. Mining associ´ation rules between sets of items in large databases. In Acm sigmod record, Vol. 22, pp. 207–216. ACM, 1993.
[3] Junqiang Liu, Jiuyong Li, Shijian Xu, and Benjamin CM Fung. Secure outsourced frequent pattern mining by fully homomorphic encryption. In International Conference on Big Data Analytics and Knowledge Discovery, pp. 70–81. Springer, 2015.
[4] 高橋卓巳, 石巻優, 山名早人. SV パッキングによる完全準同型暗号を用いた安全な委託 apriori 高速化. 第 15 回情報科学技術フォーラム F-002, 2016.
[5] Hiroki Imabayashi, Yu Ishimaki, Akira Umayabara, Hiroki Sato, and Hayato Yamana. Secure frequent pattern mining by fully homomorphic encryption with ciphertext packing. In International Workshop on Data Privacy Management, pp. 181–195. Springer, 2016.
[6] Kioxia, https://personal.kioxia.com/ja-jp/ssd/exceria-plus-nvme-ssd.html/
[7] Samsung, https://www.samsung.com/semiconductor/minisite/jp/ssd/consumer/980pro/
[8] Intel, https://www.intel.co.jp/content/www/jp/ja/products/memory-storage/solid-state-drives/consumer-ssds/optane-ssd-9-series/optane-ssd-905p-series/905p-380gb-m-2-\\110mm-20nm.html/