2ZD-02

完全準同型暗号を用いたゲノム秘匿情報検索アプリケーションにおける不揮発メモリ活用の検討

廣江　彩乃†　　小口　正人†

† お茶の水女子大学

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．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．
本研究では，近年注目が集まっている不揮発メモリを活用することを考える． 不揮発メモリは Persistent Memory とも呼ばれ，メインメモリの処理速度の速さと，SSD や HDD のストレージの長所を盛り込むことを目的に開発された．これを秘匿検索に用いて，その際の実行時間の比較・評価を行う．そして， 費用の面も交えた観点から，不揮発メモリの有効な活用方法を検討する．

2 先行研究

石巻らの (2016) 先行研究によるゲノム秘匿検索アプリケーション [3] を，本研究のメモリ性能評価に用いる．図 1 にアプリケーションの流れを示す．これはサーバとクライアントが 1:1 で問い合わせを行うものである．また，ゲノムデータは A,C,G,T の四文字から成る配列であるため，文字列検索と見なせる．サーバはクライアントから，検索したい文字列を暗号化処理したものと，その文字列を検索したい配列上の検索開始地点を受け取り，秘匿検索を行ってマッチしたか否かの結果を返す．クライアントから送られる暗号化された文字列をクエリ，文字列検索開始地点をポジションと呼ぶ．

このアプリケーションで用いるゲノム配列のデータベースは，検索の高速化のため，離散データ構造である Positional-Burrows Wheeler Transform (PBWT) [2] の形に変換している．これはゲノムデータに対して列ごとのソートを行ったもので，計算量を大幅に削減することが出来る．また，クライアントがサーバに，ダミーを含めた複数の検索開始ポジションを伝えることで，実際に利用するポジションを秘匿することが出来る等，秘匿性向上のための工夫も為されている．

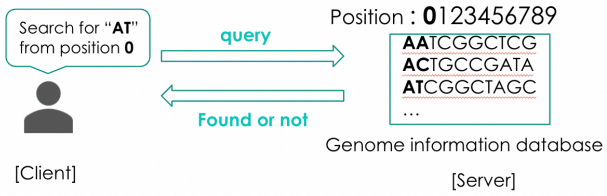

図 1　アプリケーションの流れ

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ゲノム秘匿検索アプリケーションを用いて実行時間を計測し，実行条件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 先領域に不揮発メモリを指定して，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 実行条件

swap デバイスに不揮発メモリを用いた場合の性能評価を行っていく．メモリが不足して swap 処理が発生する状況を再現するため，使用可能メモリを段階的に制限した複数の条件下で実験を行った．制限の方法は，アプリケーション実行をコンテナ内で行いそのコンテナで使用可能なリソースを設定することで制限した．条件としては，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，不揮発メモリに

Research for an Application of Genome Information Search with Fully Homomorphic Encryption on Persistent Memory Devices
†Ayano Hiroe　†Masato Oguchi
†Ochanomizu University

swap 領域を作成させる条件 (2)-(5) である．各条件で制限した使用可能メモリは表 3 に示す．

表 1　サーバ性能

CPU	Intel®Xeon®Silver 4313 16 Cores
DRAM	DDR4-2667 16GB*12=192GB 3200MHz
PCIe Gen	4.0

表 2　不揮発メモリ性能

製品名	Intel Optane 200 Series
容量	128GB *2 = 256GB
インタフェース	DDR-T

3.3　実験結果及び分析

ゲノム秘匿検索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実行時間を取得した．実行の際のコンテナ使用可能メモリと，アプリケーション実行時のパラメータであるクエリ長は表 3 に示す通りである．これらを各々組み合わせて 15 種類の条件ごとに実行時間を計測した．クエリ長とは，配列が合致するものがあるか，探索したい対象のゲノムデータの長さである．クエリ長が多くなると演算量が多くなり，swap 量が多くなる．

表 3　実 行 条 件

使用可能メモリ (5 段階)	十分 0.7GB 1GB 2GB 5GB
パラメータ (クエリ長) (3 段階)	4 5 6
使用可能不揮発メモリ (swap 領域)	十分

実行時間の結果を図 2 に，各クエリ長ごとに折れ線グラフとして示す．まず図 2 の通り，クエリ長ごとの折れ線が交わることなく推移し，クエリ長が長くなるにつれて実行時間は長くなることがわかる．次に，図 3 には，各パラメータ使用時にメインメモリを十分に与えた条件下での実行時間で水平方向に線を引き，それと同じパラメータ使用時のグラフとの交点を星マークで示した．また，表 4 に示す通り，この際の swap 量は使用可能メモリの 5 倍もの swap 領域である．これほど swap 処理が発生していても，不揮発メモリの高速な処理能力が発揮され，メインメモリ上のみで実行した時と実行時間が変わらないことがわかる．

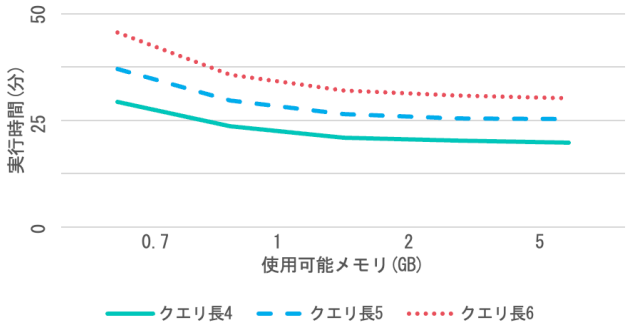

図 2　使用可能メモリと実行時間の推移

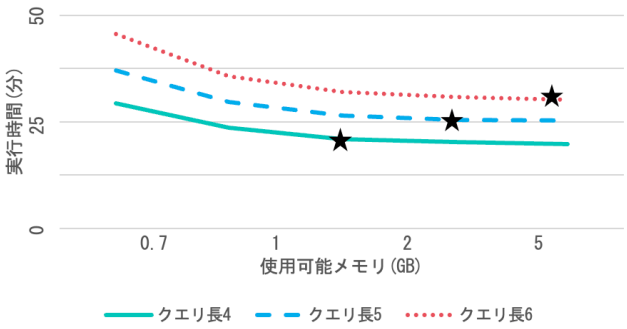

図 3　各パラメータにおいて実行時間短縮に効果があるメインメモリ量
※星マークは，メインメモリを用いることで実行時間を短縮できる限界値を表し，それより多くメインメモリを用いても実行時間短縮の効果は得られないことを意味する．

表 4　発生 swap 量の目安

使用可能メインメモリ 1GB 条件下での swap 領域最大使用量	約 8 GB

4　まとめと今後の課題

今回行った実験結果より，不揮発メモリの高速な処理能力が発揮され，使用可能メインメモリ量の 5 倍以上もの swap 領域が使われている時にも実行時間が変わらないことがわかった．また，アプリケーション実行条件 (使用パラメータ) によって実行時間の短縮に効果を発揮するメインメモリの量が異なることもわかった．従来のストレージ製品と比べて不揮発メモリの処理能力は遥かに高く，メインメモリと比べると価格は低いということを踏まえ，swap 処理が発生するようなアプリケーション実行においてメインメモリと不揮発メモリを効果的に併用することが今後の完全準同型暗号の実用化に向けて一つの鍵となると言える．

謝　　辞

本研究の一部は，キオクシア株式会社と JST CREST JPMJCR22M2 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] R. Durbin, ”Efficient haplotype matching and storage using the Positional Burrows-Wheeler Transform (PBWT),” Bioin- formatics, vol. 30, no. 9, pp. 1266-1272, 2014
[3] Y. Ishimaki et al., “Privacy-preserving string search for genome sequences with FHE bootstrapping optimization.” 2016 IEEE International Conference on Big Data (Big Data). IEEE. 2016, pp. 3989–3991.