オンラインデータの安全な利用に向けた共通鍵検索可能暗号の研究

Searchable Symmetric Encryption for Secure Utilization of Online Data

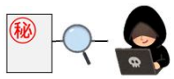
佐藤敬恒・法制倫理分科会・情報セキュリティ大学院大学

Abstract: With the spread of the online preservation, Searchable Symmetric Encryption (SSE), which realizes searching of encrypted documents through encrypted queries, has been paid attention. A lot of SSE scheme allow the leakage of access patterns and search patterns. However, recent works show that adversaries having some auxiliary information can exploit those information leakage, and perform query recovery attacks to guess keywords corresponding to encrypted search queries. Although an SSE framework proposed by Chen et al. (IEEE INFOCOM 2018) prevents access pattern leakage with use of erasure coding to make documents redundant into multiple shards and perturbing the registration of these shards to encrypted indexes, this framework doesn't provide the protection for query recovery attack based on search pattern leakage. We propose an SSE scheme to create several small indexes, which contains identifiers of document's shards obtained by Erasure Coding. Our scheme offers countermeasure of query recovery attacks based on access pattern leakage or search pattern leakage.

1.検索可能暗号とは

●クラウド活用が推進される中、**オンライン上にデータを保存する需要も拡大**しつつある。

●データを暗号化しなければ、サーバ管理者の**不正な閲覧やデータ流出による情報漏洩が心配**である。



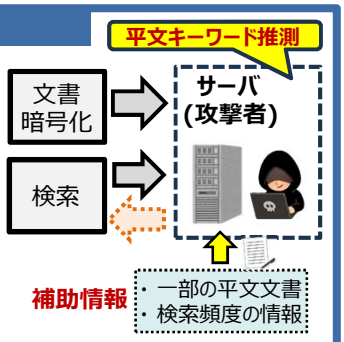
●通常の暗号化では、**大量のデータの中から必要なデータを検索することができないため不便**である。

➡ 暗号化状態で検索を実現する**検索可能暗号**が注目される。

2.クエリ回復攻撃とは

●暗号化文書の保存や検索の処理過程で、**サーバ(攻撃者)は検索時期や検索結果の応答件数を観測**できる。

●サーバが平文文書の一部や検索頻度の情報を知っていると、**クエリに対応する平文キーワードを推測**しうる。(**クエリ回復攻撃**)

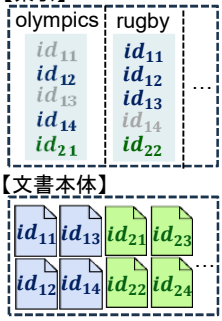


3.既存の対策方式[1]

●データ保護技術である**Erasure Coding**を使用して，登録文書を冗長化する。

●冗長化した**文書(破片文書)の一部をキーワードと関連づけて索引登録**する。

●キーワードと**無関係な破片文書もダミーとして索引へ登録**する。



※破片文書2つで文書を復元できるとする

(例) olympicsとrugbyともid_1の文書と対応しているとする。対策を適用すると**id_1の文書はid_{11}~id_{14}に冗長化**され、撹乱された索引が構成される。これによりクエリ回復攻撃を困難にする。

4.検索頻度の情報を用いた攻撃

●[1]の対策方式では、キーワードの検索時期は撹乱されず、**攻撃者は検索頻度を用いたクエリ回復攻撃**を実行しうる。

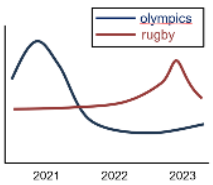


図 世間の注目状況
(イメージとして作成)

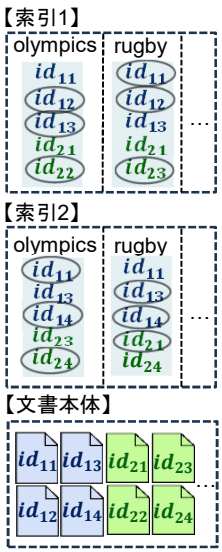
(例) クエリq_1はolympicsとrugbyのいずれかのキーワードと対応しているとする。**q_1は2021年に多く発行されたとすれば、世間の注目状況からq_1はolympicsに対応する**と推測できる。

5.提案方式の工夫

●**Erasure Coding**を使用して生成した破片文書のうち、決まった種類のものを用いた**索引を複数件用意**する。

●索引で登録対象となる破片文書のうち、**キーワードとの対応関係があるものは、全て関連付けを**する。また索引で登録対象となる破片文書のうち、**キーワードとの対応関係がないものも、ダミーとしてランダムに関連付け**る。

●検索では**キーワードとともに、任意の索引番号を選択**する。対応関係にある破片文書と、ダミーの破片文書の取得割合も考慮のうえで検索する。



(例) **破片文書を3種類ずつ用いて、2つの索引を用意**する。検索においては、対応関係にある**一定数以上の破片文書を○のように取得**する。ダミー文書の検索も行い撹乱する。

6.提案方式と検索頻度の分散

●提案方式では攻撃者はユーザの選択した索引番号を困難にする。これにより**複数の索引に検索先を適切に分散させる**ことで、観測されるクエリの発行頻度を撹乱する。

●分散方法として、索引ごとに選択確率を**重み付け確率を設定**した上、検索回数や時期に応じて**不定期に設定値を更新**するなどが考えられる。

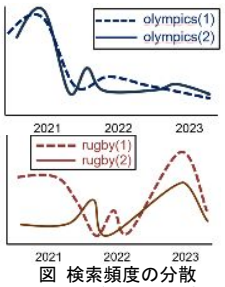


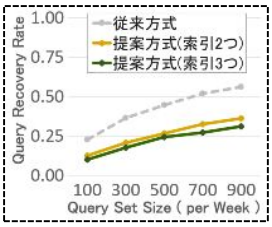
図 検索頻度の分散
(イメージとして作成)

7.提案方式に対する評価結果

●Google Trends[2]と同じ検索頻度で、ユーザがキーワードの暗号化検索を行い、**攻撃者が全クエリのうち正確に予測できた割合を評価**した。

●Enronデータセット[3]の上位**250件**のキーワードを検索対象とする。

●攻撃者は検索キーワードの種類とGoogle Trends上の**検索頻度を知っている**前提とする。



➡ **攻撃者の正解率は低下し、有効性を確認**できた。

[1] G. Chen, T. Lai, M. K. Reiter, and Y. Zhang. Differentially private access patterns for searchable symmetric encryption. In *IEEE INFOCOM 2018*, pp.810-818, 2018.
[2] Google, 2024, "Google Trends", (https://trends.google.co.jp/trends/, 2024-02-07 参照)
[3] William W. Cohen, 2015, "Enron email dataset", (https://www.cs.cmu.edu/~./enron/, 2024-02-07参照)