

Private Information Retrieval

胡瀚林

May 18, 2016

① 背景

② 守り方

③ 攻撃方

④ 参考文献

① 背景

② 守り方

③ 攻撃方

④ 参考文献

Private Information Retrieval



- Q:検索質問
- R(Q):質問 Q の検索結果

Location vs Keyword

- Location
 - 地図
 - 乗換案内
 - 近くのレストラン
- Keyword
 - ウェブ検索
 - データベース検索
 - クラウドストア検索

AOL 事件

AOL 質問ログ

AnonID	Query	QueryTime	ItemRank	ClickURL
4417749	care packages	2006-03-02 09:19:32	10	http://booksforsoldiers.com
4417749	care packages	2006-03-02 09:19:32	9	http://www.brandonblog.com
4417749	movies for dogs	2006-03-02 09:24:14		
4417749	blue book	2006-03-03 11:48:52	1	http://www.kbb.com
4417749	best dog for older owner	2006-03-06 11:48:24	1	http://www.canismajor.com
4417749	best dog for older owner	2006-03-06 11:48:24	5	http://dogs.about.com

- 2006年8月4日、AOL(American OnLine)が650,000人以上のユーザーの匿名化された検索質問ログを研究目的でリリースした。

AOL 事件

AOL 質問ログ

AnonID	Query	QueryTime	ItemRank	ClickURL
4417749	care packages	2006-03-02 09:19:32	10	http://booksforsoldiers.com
4417749	care packages	2006-03-02 09:19:32	9	http://www.brandanblog.com
4417749	movies for dogs	2006-03-02 09:24:14		
4417749	blue book	2006-03-03 11:48:52	1	http://www.kbb.com
4417749	best dog for older owner	2006-03-06 11:48:24	1	http://www.canismajor.com
4417749	best dog for older owner	2006-03-06 11:48:24	5	http://dogs.about.com

- 2006 年 8 月 4 日、AOL(American OnLine) が 650,000 人以上のユーザーの匿名化された検索質問ログを研究目的でリリースした。
- 2006 年 8 月 9 日、ID 4417749 の名前、年齢、住所などが特定された。(?)

Location vs Keyword



猫 ? 犬

- 位置間の距離は簡単に計算できるが、単語間の距離は計算しにくい
- 単語の次元数が高い



猫 ?

- ノイズを加えにくい

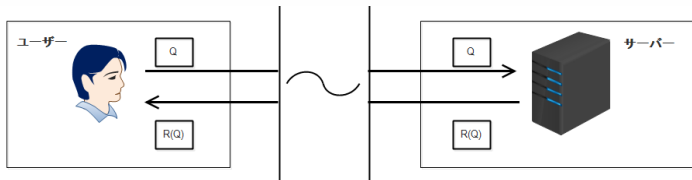
① 背景

② 守り方

③ 攻撃方

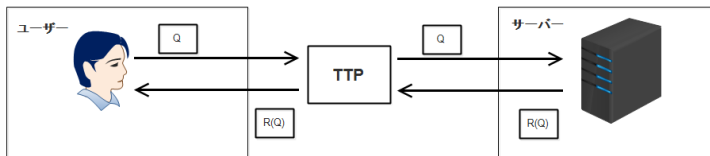
④ 参考文献

Anonymity



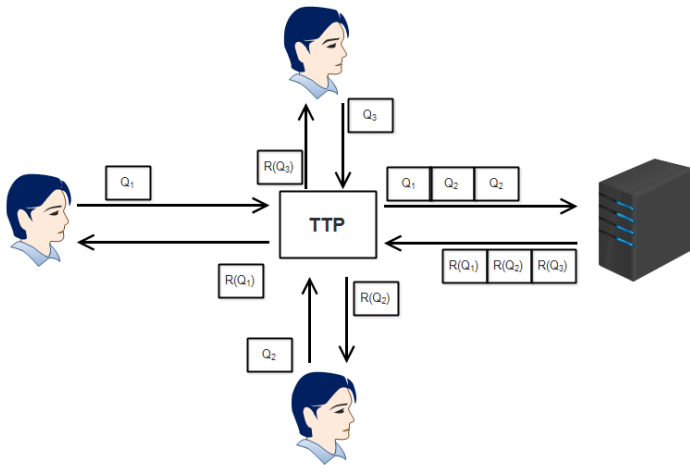
- 質問者を隠す

Tursted Third Party



- 質問者の IP アドレスなどを隠す

Tursted Third Party



- 複数の質問者を混ぜて検索する

Perturbation

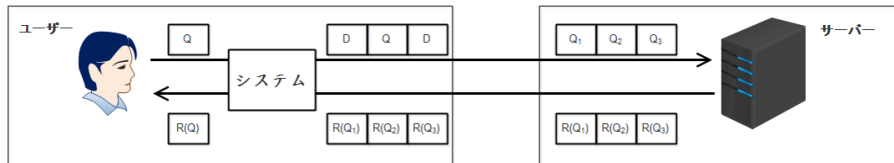
Location

- Geo-indistinguishability (?)

Keyword

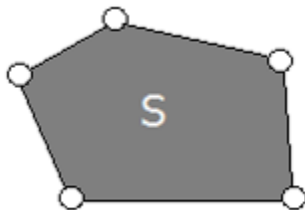
- 質問を一般化して検索する (?)
リンゴ ⇒ 赤 果物
- 事前に標準質問を作って、本当の質問の代わりに使う (?)

Obfuscation



- 複数の質問を混ぜて検索する

Obfuscation-Location



定義 $((k, s) - \text{privacy} (?))$

本当の位置と $k - 1$ 個のダミー位置に囲まれた図形の面積が S 以上ある

Obfuscation-Keyword (?)

問題

どのようなダミー質問がいいダミー質問

Obfuscation-Keyword (?)

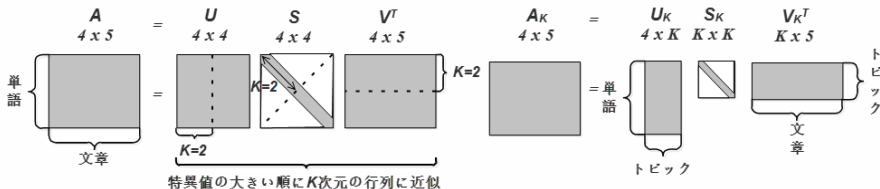
問題

どのようなダミー質問がいいダミー質問

Plausibly Deniable Search (?)

- 本当の質問との“ 距離 ”が遠い
- 本当の質問と似たような“ 確率 ”で提出される

Latent Semantic Analysis



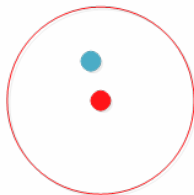
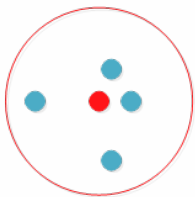
潜在的意味インデキシング

単語・文書行列 A を特異値分解 $A = USV^T$ し、 U 、 S 、 V の各列ベクトルを特異値が大きい順に K 個用いて A の低ランク近似 $A_K = U_K S_K V_K^T$ を得る。

このように低ランク分解によって、単語とトピックの関係を分析することができる

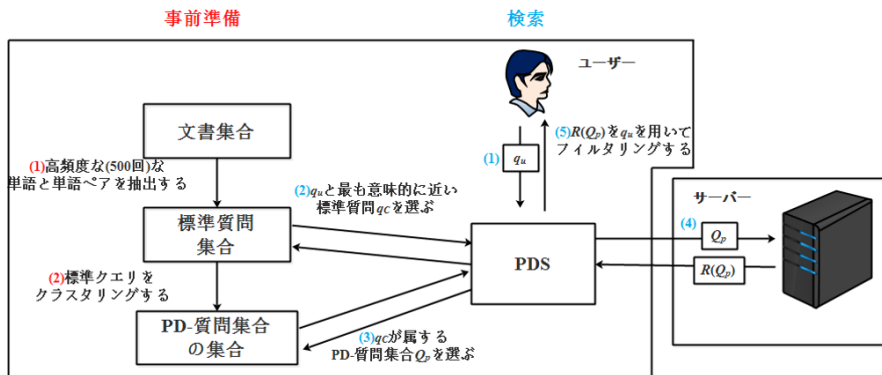
Plausibly

- 標準質問
- 質問ログ

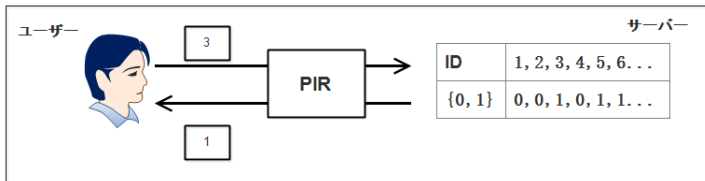


- 質問の近傍の中の質問数で“ 確率 ”、あるいは尤もらしさを計算する
- 質問数が多いほど“ 確率 ”が高いとする

Plausibly Deniable Search



PIR (?)



- 暗号などの手法を用いて質問の内容を完全に隠す

準同型暗号

定義 (準同型暗号)

二つの暗号文 $Enc(m_1)$, $Enc(m_2)$ が与えられた時に、平文や秘密鍵なしで $Enc(m_1 \circ m_2)$ を計算できる暗号

例 (加算ができる準同型暗号)

$Enc(\cdot)$: 暗号化 $Dec(\cdot)$: 復号

$$Dec(Enc(m_1) \cdot Enc(m_2)) = m_1 + m_2$$

準同型暗号

ユーザー

質問生成

```
1: Input:  $i^*, n$ 
2: for  $i = 1, \dots, n$  :
3:   if  $i == i^*$  :
4:      $q_i = \text{Enc}(1)$ 
5:   else
6:      $q_i = \text{Enc}(0)$ 
7: return
    $Q = \{q_1, \dots, q_n\}$ 
```

復号

```
1: input:  $R$ 
2: return  $\text{Dec}(R)$ 
```

サーバー

結果計算

```
1: Input:  $Q, \{x_1, \dots, x_n\}$ 
2:  $R = 0$ 
3: for  $i = 1, \dots, n$  :
4:    $R = R \cdot q_i^{x_i}$ 
5: return  $R$ 
```

Note

$m_1 = m_2 \nRightarrow \text{Enc}(m_1) = \text{Enc}(m_2)$
 $\text{Dec}(R) = \sum_{x_i=1} \text{Dec}(q_i) = x_{i^*}$

PIR I

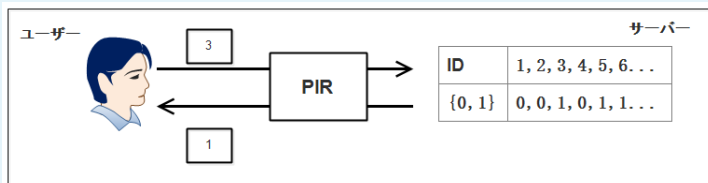
- 1995 Chor et al. : Multiserver PIR
 - 情報理論から見ると single-database PIR ができない
- 1997 Kushilevitz and Ostrovsky : computational single-database PIR
 - *quadratic residuosity computational assumption*
 - 通信量: $O(2^{\sqrt{\log n \log \log N}})$
- 1999 Cachin et al. : s-PIR
 - Φ – *hiding number – theoretic assumption*
 - 通信量: $O(\log^8 n)$
- 2000 Kushilevitz and Ostrovsky : Private Block Retrieval
 - *Naor – Yung one – way 2 – to – 1 trapdoor permutations*
 - *Goldreich – Levinhard – corepredicates*
 - 通信量: $n - cn/2k + O(k^2)$

PIR II

- 2005 Gentry and Ramzan : Multiserver PIR
 - Φ – *hiding number* – *theoretic assumption*
 - 通信量: $O(\log^2 n)$
- 2007 Aguilar-Melchor and Gaborit : computationally-efficient PIR
 - *lattice – based*
 - a few thousand bit-operations per bit in the database
 - 2010 Olumofin and Goldberg: 応答時間は普通の方法の千分の一くらい
- 2013 Yi et al. : PBR
 - *Fully homomorphic encryption*
 - 通信量: $(\gamma + \gamma/)$
 - 計算量: $(\gamma^2 + \gamma/2)$
 - 計算時間: 2min
 - 通信時間: 4.5s (100 – Mb/second)

Obfuscation + PIR

Embellishing Text Search Queries to Protect User Privacy



質問ではなく単語ごとにダミーを加える

Obfuscation + PIR

Query processing by the search engine

```
1: Input: Embellished query  $q$ 
2: Let  $R = \phi$ 
3: for all  $\langle t_i, E(u_i) \rangle \in q$  :
4:   for all  $\langle d_j, p_{ij} \rangle \in L_i$  :
5:     if  $\exists \langle d_j, E(score_j) \rangle \in R$  :
6:        $E(score_j) = E(score_j) \cdot E(u_i)^{p_{ij}}$ 
7:     else
8:       Insert  $\langle d_j, E(u_i)^{p_{ij}} \rangle$  into  $R$ 
9: return  $R$ 
```

① 背景

② 守り方

③ 攻撃方

④ 参考文献

① 背景

② 守り方

③ 攻撃方

④ 参考文献

Bibliography I