秘密分散データを用いた医療データ検索システム

1235067　中村　巴　【 ネットワーク信号処理研究室 】

Medical Record Retrieval System Using Secret Sharing Data

1235067　NAKAMURA, Tomo　【 Signal Processing and New Generation Network Lab. 】

1　はじめに

広域災害に備え医療データを保全し，被災地での診療に活用することが求められている．災害時は，ネットワークや電源などのリソース不足が考えられるため，診療に最低限必要な情報のみを入手できると良い．医療データの保全に適した方法の 1 つとして (k,n) しきい値秘密分散法 [1] を用いた分散バックアップがあるが，分散したデータの一部だけを部分的に復元することはできない．そこで秘密分散したデータの一部分だけを復元する部分復元可能な秘密分散法が提案された [2]．分散した医療データの検索の仕組みは無く，名前などの患者を識別するための情報を部分的に復元してから検索する方法が考えられる．復元は方程式を解くための逆行列の作成にかかる計算量が大きく，時間を要することが想定される．本研究では安全性を落とすことなく検索にかかる計算量の削減を目的として，識別情報を復元することなくシェアの状態で検索を可能とする方法を提案する．

2　部分復元可能な秘密分散法

秘密分散したデータから一部のデータを部分的に復元する方法が提案された．この方法は，分散したいデータを意味のある項目ごとに分割し，分割した項目それぞれに対して秘密分散する．そして復元したい項目のシェアのみを結合することで部分的な復元を実現している．分散したデータの検索の仕組みは無く，図 1 に示すように，名前など患者を識別するための情報を部分的に復元してから検索する方法が考えられる．復元には方程式を解く必要があり，バックアップしている全員分の識別情報を部分復元してから検索を行うため時間を要することが想定される．そのため，医師は診療を始めることができないという問題がある．

3　シェア間の係数差の比較による検索

検索の高速化を目的として，識別情報を復元することなくシェアの状態で検索を可能とする方法を提案する．本節ではシェアの状態で検索を可能とする方法について述べ，シェアとタグの保存方法について述べる．

図 1　部分復元可能な秘密分散法の検索の流れ

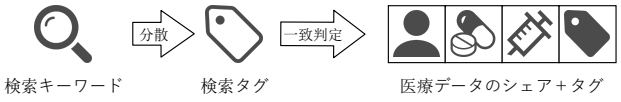

図 2　シェア間の係数の差の比較による検索の流れ

3.1　シェアの状態で検索を可能とする方法

提案方法の検索の流れを図 2 に示す．提案方法は識別情報 kw を $(2,n)$ しきい値秘密分散してタグ t_u $(i = 1,2,\cdots,n)$ を作成する．タグは式 (1) のような合同式で表される．

$$t_u \equiv kw + rx_u \pmod{p} \tag{1}$$

タグ t_u を医療データのシェアに対応付けて保存し，タグの作成に使用した x_u，p も同時に保持する．

検索時は保持している x_u，p を用いて検索キーワード kw' を $(2,n)$ しきい値秘密分散して検索タグ t'_u を作成する．検索タグは式 (2) のような方程式で表される．

$$t'_u \equiv kw' + r'x_u \pmod{p} \tag{2}$$

ここで，t_u と t'_u の添字が同じもの同士の減算を行い，その結果に x_u の逆元 x_u^{-1} を乗算すると，

$$(t_u - t'_u)x_u^{-1} \equiv (kw - kw')x_u^{-1} + (r - r') \pmod{p} \tag{3}$$

となる．式 (3) より $kw = kw'$ の場合係数の差 $r - r'$ を正しく求めることができ，全ての u でこの結果は同じ値になる．一方 $kw \neq kw'$ の場合全ての u でこの結果が同じ値になることはない．また，n 個すべて比較する必要はなく，2 個比較すれば一致判定が可能である．これを利用してシェアの状態で検索を可能としている．

この方法により，従来法では一致判定に $\mathcal{O}(l^3)$(医療データのデータ長を l とする) の計算量がかかっていたところ，提案法では $\mathcal{O}(d^2)$(識別情報のデータ長を d とする) の計算量となる．

3.2　シェアとタグの保存方法

タグと検索タグの係数の差を求めるには，添字が同じもの同士を正しく比較する必要がある．そのためには同一の識別情報から作成したタグの組み合わせを知っている必要があるが，タグと医療データのシェアは対応付けて保存しており，医療データのシェアの組み合わせも分

表 1　添字が 1 のストレージ

医療データのシェア	タグ $_1$	タグ $_2$
A さんシェア $_1$	A さん t_1	A さん t_2
B さんシェア $_1$	B さん t_1	B さん t_2
⋮	⋮	⋮

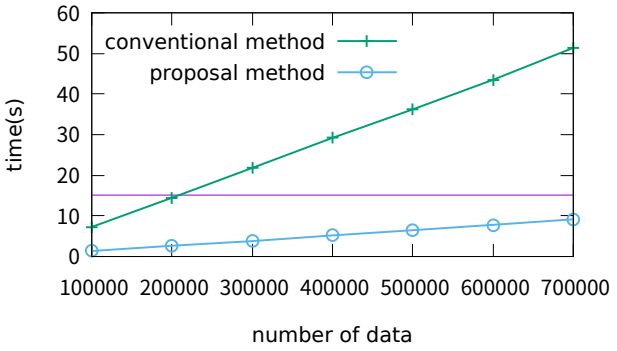

図 3　従来方法と提案方法の検索時間の比較

かるため医療データを不正に復元されてしまう恐れがある．そのため，タグの組み合わせは分かるが医療データのシェアの組み合わせが分からないような仕組みが必要となる．タグは 2 個あれば一致判定が可能であるため，表 1 のように 1 個のシェアに対しタグを 2 個ずつ付与することで各ストレージで一致判定をすることが可能である．これにより医療データのシェアとタグの組は各ストレージでランダムに保存することが可能となる．その結果，2 個以上の医療データのシェアの対応を知ることはできず，同様に 3 個以上のタグの対応を知ることはできない．

4　評価

識別情報を復元してから検索する従来の方法と提案方法の検索時間を比較し，安全性について評価する．

4.1　検索時間の比較

識別情報と検索キーワードを 144bit，1 つのデータを 3000byte とし，検索キーワードを入力してから該当するデータを出力するのにかかった時間を chrono 関数を用いて計測した．結果を図 3 に示す．データ数分同じ処理を繰り返しているため，どちらも線形となっている．また，同じ時間内に検索することができるデータ数の差は時間に比例して増加していくと考えられる．

4.2　安全性の評価

提案方法では，n 個あるタグのうち 2 つ t_1, t_2，k と x_1, x_2 あるいは p が漏洩した場合には，kw を有限個に絞り込まれる．この場合について p を大きくすれば問題ないことが分かっているが，識別情報を絞られることによって万が一医療データを不正に復元されてはならないことから，t_u，x_u，p をシステムで秘匿することで安全性を担保する．

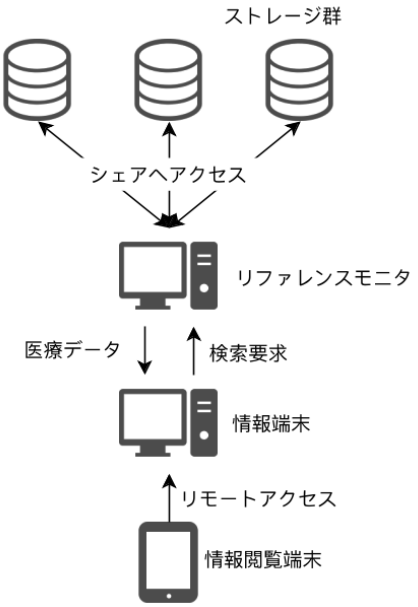

図 4　検索システムの構成

5　検索システム

システム構成を図 4 に示す．医療データのシェアとタグを保存するストレージにアクセスできる端末を制限するため，検索・復元用の端末を拠点病院内に設置しこれをリファレンスモニタと呼ぶ．また，リファレンスモニタに検索要求を送信できる端末を制限するため，検索要求送信用の端末を拠点病院内に設置しこれを情報端末と呼ぶ．医師が使用する端末には盗難などの危険性があるためいかなる情報を保持しておくのは好ましくない．そのため，医師は情報閲覧端末から情報端末にリモートアクセスすることで医療データの検索・閲覧を行う．このシステム構成により安全性を担保できると考える．

6　まとめ

本研究では安全性を落とすことなく検索にかかる計算量の削減を目的として，識別情報を復元することなくシェアの状態で可能とする方法を提案した．計算量は削減できたが，n 個あるタグのうち 2 つ t_1, t_2，k と x_1, x_2 あるいは p が漏洩した場合には識別情報 kw が有限個に絞り込まれる．これらの場合に対して提案したシステム構成を用いることで安全性を担保できると考える．

参考文献

[1] A. Shamir, “How to Share a Secret,” Communication of ACM, Vol. 22, No. 11, pp. 612–613, Nov. 1979.

[2] 田中麻実，福冨英次，福本昌弘，“秘密分散バックアップした医療データの部分復元,”信学技報 IA2015-74，pp. 31–36，Dec. 2015.