令和 2 年度

修士学位論文

秘密分散データを用いた医療データ検索システム

Medical Record Retrieval System Using Secret Sharing Data

1235067　　中村　巴

指導教員　　福本昌弘

2021 年 3 月 5 日

高知工科大学大学院 工学研究科 基盤工学専攻

情報学コース

要 旨

秘密分散データを用いた医療データ検索システム

中村　巴

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．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．

本論文では安全性を落とさず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している．提案方法の検索時間を計測した結果 70 万人分のデータを 15 秒以内に検索できることを確認している．また，提案検索法を用いたシェアの保存方法により第三者がストレージにアクセスしても医療データを不正に復元される恐れはない．しかし，ある条件下において識別情報が漏洩する場合があり，有限体の法を大きくすれば問題ないが医療データを扱うには十分であるとは言えず対策を講じたシステムを提案している．提案システムを用いることで識別情報の漏洩を防止可能であり，安全に医療データの検索を行うことができる．

キーワード　秘密分散法，部分復元，検索

Abstract

Medical Record Retrieval System Using Secret Sharing Data

Tomo NAKAMURA

It is required to preserve the medical record in preparation for wide-area disasters and utilize it for medical treatment in the disaster area. In the event of a disaster, resources such as networks and power supplies may be insufficient, so it is good to be able to obtain only the minimum information necessary for medical treatment. One of the methods suitable for the preservation of medical records is distributed backup using the (k, n) threshold secret sharing scheme, but it is not possible to partially decode only a part of the distributed data. Therefore, a confined decodable secret sharing scheme was proposed that decodes only a part of the secretly shared data. There is no mechanism for retrieve distributed medical records, and a method of retrieving after partially decoding information for identifying a patient, such as a name, is conceivable. Decoding is expected to require a large amount of calculation and time to create the inverse matrix for solving the equation.

In this paper, we have proposed a method that enables retrieval in a shared state without decoding the identification information, to reduce the amount of calculation required for retrieval without compromising safety. As a result of measuring the retrieval time of the proposed method, it has been confirmed that the data for 700,000 people can be retrieved within 15 seconds. Also, there is no risk of unauthorized decoding of medical records even if a third party accesses the storage by the share storage method using the proposed retrieval method. However, the identification information may be leaked under certain conditions, and although there is no problem if the finite field

method is increased, it is not sufficient to handle medical records, and we have proposed a system with countermeasures. By using the proposed system, leakage of identification information can be prevented, and medical records can be searched safely.

key words secret sharing scheme, confined decoding, retrieval

目次

目次

図目次

表目次

第 1 章

はじめに

1.1 本研究の背景と目的

東日本大震災では，津波によって沿岸部の病院から医療データが消失し，被災地での医療行為に支障が出た．これより，医療データを保全し災害時に活用することが求められている [1]．ただ保全をすれば活用できるわけではなく，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，全ての医療データを通信するには不安が残る．また，医師によると災害時における適切な診療には投薬歴やアレルギー情報などのデータがあれば必ずしも全ての医療データが必要とは限らない．したがって，災害時に活用するためには診療に最低限必要な情報のみを入手できる仕組みが必要となる．保全方法として，医療データを遠隔地にバックアップする方法が考えられる．バックアップは真正性，見読性，保存性，冗長性，秘匿性を確保する必要があり，適した手法として (k, n) しきい値秘密分散法 [2] を用いた分散バックアップがある．しかし，(k, n) しきい値秘密分散法は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，秘密分散バックアップした医療データの部分復元可能な秘密分散法が提案された [3]．このシステムは，(k, n) しきい値秘密分散したデータの一部分だけを復元することができる．しかし，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．識別情報を部分的に復元するためには，同一人物の分散データの組み合わせを知っている必要があり，対応づけてバックアップしておく必要がある．しかし，第三者がバックアップデータにアクセス可能であれば分散データの組み合わせが分かり，医療データを不正に復元される恐れがある．また，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，多数の被災者を素速く診療する必要がある災害時にはこの方法は適していない．

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，患者の識別情報を復元すること無くシェアの状態で検索を実現する方法を提案する．医師によると 15 秒程度で医療データが手元に欲しいことから，高知県の人口である 70 万人分のデータを 15 秒以内に検索することを目標とする [4][5]．さらに，第三者にバックアップデータにアクセスされた場合でも，医療データの不正復元を防止する仕組みを提案する．そして，識別情報を復元してから検索する従来検索方法と提案検索方法の比較を行い，提案検索法の安全性の評価を行う．最後に，提案検索方法を用いた医療データの検索システムについて述べる．

1.2　本論文の構成

本節では本論文の構成について述べる．2 章ではバックアップの対象となる医療データについて述べ，バックアップ手法である (k, n) しきい値秘密分散法について述べる．3 章では部分復元可能な秘密分散法について述べ，その検索方法を述べる．4 章ではシェアの状態で検索を可能とする方法を提案し，シェアとタグの保存方法について検討する．5 章では提案方法の検索時間と安全性について評価を行う．6 章では検索システムの構成について述べる．7 章では本研究をまとめ，今後の課題を述べる．

第 2 章

医療データのバックアップと災害時での利用

東日本大震災では，津波によって沿岸部の病院に保存していたカルテなどの医療データが消失した．被災地では，処置を行うために患者が服用していた薬などの情報が必要となり患者に聞き取りを行ったが，正確な情報を手に入れることができず適切な診療を行えないなどの支障が出た．これを受け医療データを保全して災害時に活用することが求められている．医療データを保全するには分散バックアップが有効であり，適している手法として (k, n) しきい値秘密分散法がある．

本章では，バックアップの対象となる医療データについて述べ，バックアップの条件について述べる．また，バックアップ手法である (k, n) しきい値秘密分散法について述べる．

2.1 バックアップ対象となる医療データ

医療データのうち，本研究で対象とする電子カルテについて述べ，電子カルテの標準規格である厚生労働省電子的診療情報交換推進事業 (SS-MIX: Standardized Structured Medical Information eXchange) について述べる．

2.1.1 電子カルテ

電子カルテとは，医師が作成した診療記録であるカルテを電子化したものを指す．カルテを電子化することによって，保存場所の減少，検索機能の迅速化やスタッフ間の情報共有の

促進などの利点がある [6]．一方，不正アクセスなどによる情報漏洩や電子機器の故障によるカルテの破損などの欠点も存在する．

カルテを電子化する際の要件として，真正性，見読性および保存性 (電子保存の 3 原則) を満たし，最低 5 年間保存する必要がある [7]．真正性，見読性および保存性について以下に詳細を述べる．

- 真正性

  医師が作成した電子カルテに対し，保存すべき期間において故意または過失による虚偽の入力，書き換え，消去および混同が防止されており，第三者から見て作成の責任の所在が明確であること．
- 見読性

  電子カルテの内容を診療などの必要に応じて，それぞれの目的に対し支障のない応答時間や操作方法で，肉眼で見読可能な状態にでき，書面に表示できること．
- 保存性

  電子カルテを法令等で定められた期間にわたって真正性を保ち，見読可能な状態で保存されること．

2.1.2　SS-MIX

電子カルテは，電子カルテシステムを提供するベンダやシステムのバージョンにより異なるフォーマットが使用されている．そのため，異なるベンダの電子カルテシステム間での情報共有は困難であり，地域医療連携など異なる病院間でのカルテの相互参照などの妨げとなった．そこで厚生労働省は，全ての医療機関を対象とした医療情報の交換・共有による医療の質の向上を目的として厚生労働省電子的診療情報交換推進事業を開始した [8]．

SS-MIX では，医師が作成した電子カルテを HL7Ver2.5 形式に変換し，標準化ストレージへ保存する．標準化ストレージでは，図 2.1 に示すように，定められたディレクトリ構造に従ってデータを保存する．また，HL7Ver2.5 形式に変換された患者データの例を図 2.2 に

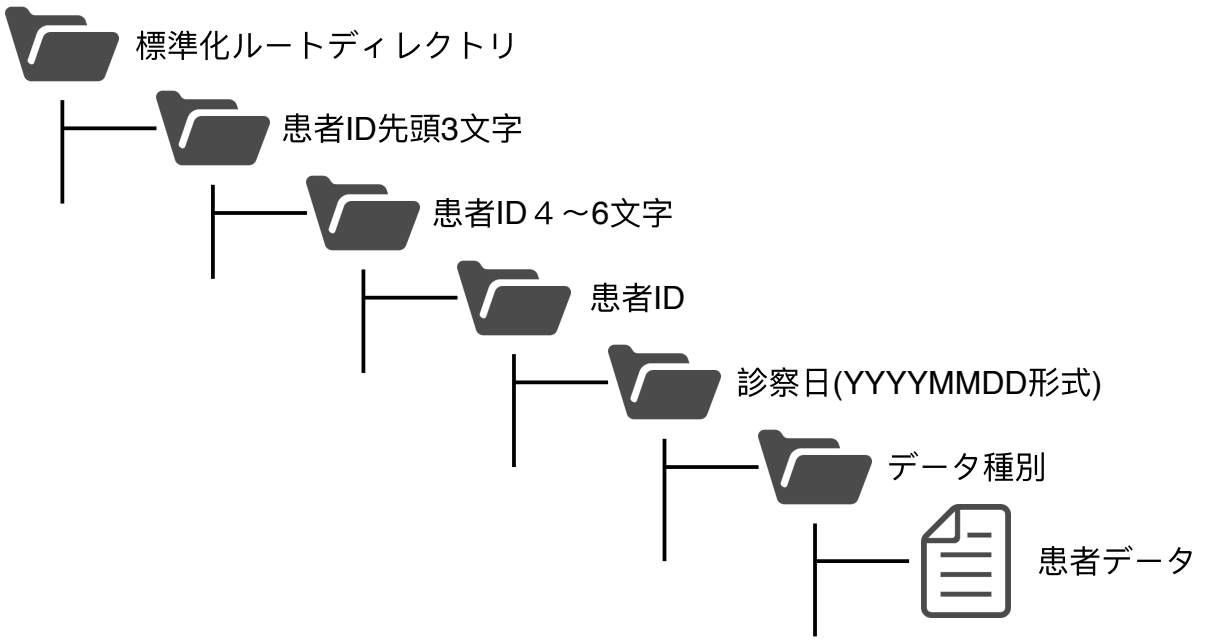

図 2.1　SS-MIX のディレクトリ構造

```
MSH  |^~¥&|EGMAIN-GX|HIS|GW|GW|20131128192616.6001||OMP^O09|20131128192616751|P|2.5|||||~ISO IR87|
     |ISO 2022-1994
ZGW  |9930000036|20131128192559|OMP-01^処方オーダ^L|13Y283154232200|INS|01^総合診療科^L
SFT  |Fujitsu Ltd.^D|V2|EGMAIN-GX
NTE  |0001|P|資源配付確認テスト
PID  |0001||9930000036||テスト^資源配布^^^^^L^I~テスト^シゲンハイフ^^^^^L^P||19750505|M||
     |高知県南国市後免町3丁目4-10^^^^783-0011^JPN^H|39|||||||||||||||N|||20130730150300|高知医療センター
PV1  |0001|O|01^^^^^C|||||01
ZIN  |01^政府管掌健康保険^HL7-0072
ZIN  |12^生活保護法による医療扶助^HL7-0072
ORC  |NW|13Y283154232200||1|||||20131128192608|KEN00100^検証^医師^^^^^^^L^^^^^I~ケンショウ^イシ^^^^^^^L^^^^^P|
     |KEN00100^検証^医師^^^^^^^L^^^^^I~ケンショウ^イシ^^^^^^^L^^^^^P|01^^^^^C|||01^総合診療科^L|MC1D016D^^L|||高知医療センター||
     |||||O^外来患者オーダ^HL7-0482
TQ1  |0001||&&&&1日1回　朝食前||||2013112800
RXO  |^リキスミア皮下注　300 μ g^MDCHOT^I1022100^リキスミア皮下注　300 μ g^L|1||KT^本^L|
     |TOC^OHP^MR9P^^外来処方~TOC^OHO^MR9P^^院外処方|||||1|KT^本^L|||Y||||||1^KT&本&L
RXR  |AP^外用^HL7-0162
```

図 2.2　標準化された患者データの例

示す.

2.2　医療データの保全方法

広域災害に備えて医療データを保全するには，遠隔地に安全にバックアップする方法が考えられる．医療データは個人情報であるため，第三者にバックアップデータから情報を読み取られないようにする必要があり，医療データをバックアップしているストレージが破損し

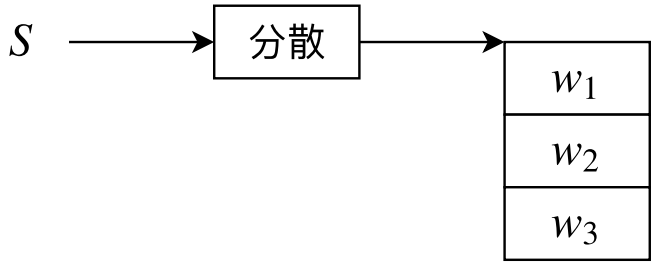

図 2.3 (k, n) しきい値秘密分散法の分散段階のデータの流れ

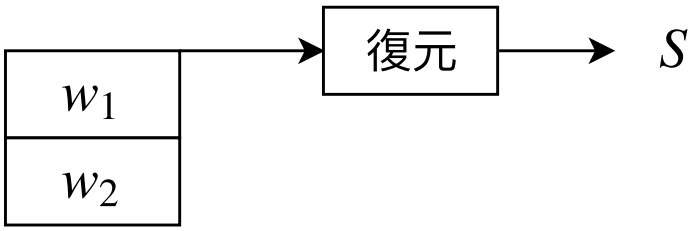

図 2.4 (k, n) しきい値秘密分散法の復元段階のデータの流れ

た場合にも，元のデータを取り出せるようにする必要がある．したがって，医療データをバックアップする際には，秘匿性，冗長性と電子保存の 3 原則を満たしている必要がある．冗長性を確保するには複数ヶ所のストレージにコピーを保存しておく方法が考えられるが，情報漏えいのリスクが高くなる．また，医療データのような個人情報を扱うには，時間をかければ必ず解けてしまう暗号では十分とは言えず，これらの条件を満たした手法として (k, n) しきい値秘密分散法を用いた分散バックアップが挙げられる．また，愛媛大学を中心とした研究グループでは秘密分散法を用いて電子カルテをバックアップする研究がある [9]．

2.3 (k, n) しきい値秘密分散法

(k, n) しきい値秘密分散法は，データを n 個のシェアとして分散し，k 個以上のシェアを集めることで元のデータを復元可能な方法である．(k, n) しきい値秘密分散法の分散段階のデータの流れを図 2.3 に示し，復元段階のデータの流れを図 2.4 に示す． 以下に (k, n) しき

2.3 (k, n) しきい値秘密分散法

い値秘密分散法の手順を示す．

分散したいデータを S，素数を $p(S < p$ かつ $n < p)$，$\mathbb{Z}/p\mathbb{Z} - \{0\}$ 上の集合を X，$\mathbb{Z}/p\mathbb{Z}$ 上の乱数の集合を R とする．すなわち，X は

$$X = \{x_1, x_2, \cdots, x_n\} (i \neq j \text{ のとき } x_i \neq x_j)$$

とおくことができる．X より

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{k-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{k-1} \end{pmatrix} \pmod{p}$$

となるような vandermonde 行列 $\mathbf{X}$ を作成する．S と R から $k \times 1$ のベクトル $\mathbf{A}$

$$\mathbf{A} = \begin{pmatrix} S \\ r_1 \\ r_2 \\ \vdots \\ r_{k-1} \end{pmatrix} (r_1, r_2, \cdots, r_{k-1} \in R, r_{k-1} \neq 0)$$

を作成する．p を法とする $\mathbf{X}$ と $\mathbf{A}$ の乗算より

$$\mathbf{XA} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} \pmod{p} \tag{2.1}$$

となるような $w_i (i = 1, 2, \cdots, n)$ が得られる．w_i をシェアと呼び，w_i と x_i を対応付けて i のストレージに分散保存する．

復元時は，シェアを k 個以上集める．集めたシェアから k 個選択し，シェアベクトル $\mathbf{W}$

$$\mathbf{W} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{pmatrix} \pmod{p}$$

を作成する．選択したシェアに対応する x_i より，

$$\mathbf{X}' = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{k-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_k & x_k^2 & \dots & x_k^{k-1} \end{pmatrix} \pmod{p}$$

となるような $k \times k$ の vandermonde 行列 $\mathbf{X}'$ を作成する．$\mathbf{X}'$ の逆行列 $\mathbf{X}'^{-1}$ を $\mathbf{W}$ の左からかけると

$$\begin{aligned} \mathbf{X}'^{-1}\mathbf{W} &= \mathbf{X}'^{-1} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{pmatrix} \\ &= \mathbf{X}'^{-1}\mathbf{X}' \begin{pmatrix} S \\ r_1 \\ r_2 \\ \vdots \\ r_{k-1} \end{pmatrix} \\ &= \begin{pmatrix} S \\ r_1 \\ r_2 \\ \vdots \\ r_{k-1} \end{pmatrix} \end{aligned}$$

となり，データ S を復元することができる．式 (2.1) を連立方程式で表すと，

$$\begin{cases} w_1 \equiv S + r_1 x_1 + r_2 x_1^2 + \cdots + r_{k-1} x_1^{k-1} \pmod{p} \\ w_2 \equiv S + r_1 x_2 + r_2 x_2^2 + \cdots + r_{k-1} x_2^{k-1} \pmod{p} \\ \vdots \\ w_k \equiv S + r_1 x_k + r_2 x_k^2 + \cdots + r_{k-1} x_k^{k-1} \pmod{p} \end{cases} \tag{2.2}$$

となり，式 (2.2) を解くとデータ S を求めることが出来る．w_i が k 個以上の場合は S を一意に定めることが出来るが，k 個未満の場合は S が一意に定まらず S を求めることが出来ない．また，n 個のシェアのうち k 個さえあれば元のデータを復元できるという冗長性と，

k 個未満のシェアでは元のデータを復元できないという秘匿性を持っている．このことから，広域災害などによって $n-k$ 個のシェアが消失しても元のデータを復元でき，$k-1$ 個のシェアからは元のデータを復元できないため，(k,n) しきい値秘密分散法は個人情報を含む医療データのバックアップ手法として適しているといえる．

しかし，秘密分散バックアップした医療データをそのまま災害時に活用できるわけでは無い．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，医療データの全てを通信するには不安が残る．また，医師によると処置を行うには投薬歴とアレルギー情報さえあれば良く，必ずしも医療データの全てが必要とは限らない．したがって，(k,n) しきい値秘密分散したデータから最低限必要な情報だけを入手できると良い．しかし，式 (2.2) から分かるように，(k,n) しきい値秘密分散法は分散したデータを復元するかしないかのどちらかであり，一部の情報だけを部分的に復元することはできない．したがって，災害時に活用するには秘密分散したデータから一部の情報だけを部分的に復元する仕組みが必要となる．

2.4　まとめ

東日本大震災では，津波によって沿岸部の病院に保存していたカルテなどの医療データが消失した．被災地では，処置を行うために患者が服用していた薬などの情報が必要となり患者に聞き取りを行ったが，正確な情報を手に入れることができず適切な診療を行えないなどの支障が出た．これを受け医療データを保全して災害時に活用することが求められている．医療データを保全するには分散バックアップが有効であり，適している手法として (k,n) しきい値秘密分散法がある．

本章では，バックアップの対象となる医療データについて述べ，バックアップの条件について述べた．また，バックアップ手法である (k,n) しきい値秘密分散法について述べた．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，医療データの全てを通信するには不安が残る．また，医師によると処置を行うには投薬歴とアレルギー情報さえあれば良く，必ずしも医療データの全てが必要とは限らない．したがって (k,n) しきい値秘密分散し

たデータから最低限必要な情報だけを入手できると良い．しかし，(k, n) しきい値秘密分散法は分散したデータを復元するかしないかのどちらかであり，一部の情報を部分的に復元することはできない．災害時に活用するには秘密分散したデータから一部の情報だけを部分的に復元する仕組みが必要となる．

第 3 章

部分復元可能な秘密分散法

医療データを保全するための方法の 1 つとして (k, n) しきい値秘密分散法を用いた分散バックアップがある．しかし秘密分散バックアップした医療データを災害時に活用するには，分散したデータのうち一部の情報のみを復元する方法を実現する必要がある．そこで秘密分散したデータの一部だけを部分的に復元する部分復元可能な秘密分散法が提案された．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秘密分散を行う．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．検索の仕組みは無く，名前など患者を識別するための情報を部分復元してから検索する方法が考えられるが，検索に時間を要するため災害時には適さない．

本章では，部分復元可能な秘密分散法について分散段階と復元段階に分けて述べ，部分復元可能な秘密分散バックアップした医療データの検索方法について述べる．

3.1 分散段階

分散段階のデータの流れを図 3.1 に示す．分散したいデータを S とする．S を意味のある項目ごとに分割したデータを

$$S_i (i = 1, 2, \cdots, d)$$

とし，S_i のデータ長を l_i とする．データ S は分割データ S_i を用いて

$$S = S_1 \prod_{m=2}^{d} 2^{l_m} + S_2 \prod_{m=3}^{d} 2^{l_m} + \cdots + S_{d-1} 2^{l_d} + S_d \tag{3.1}$$

3.1　分散段階

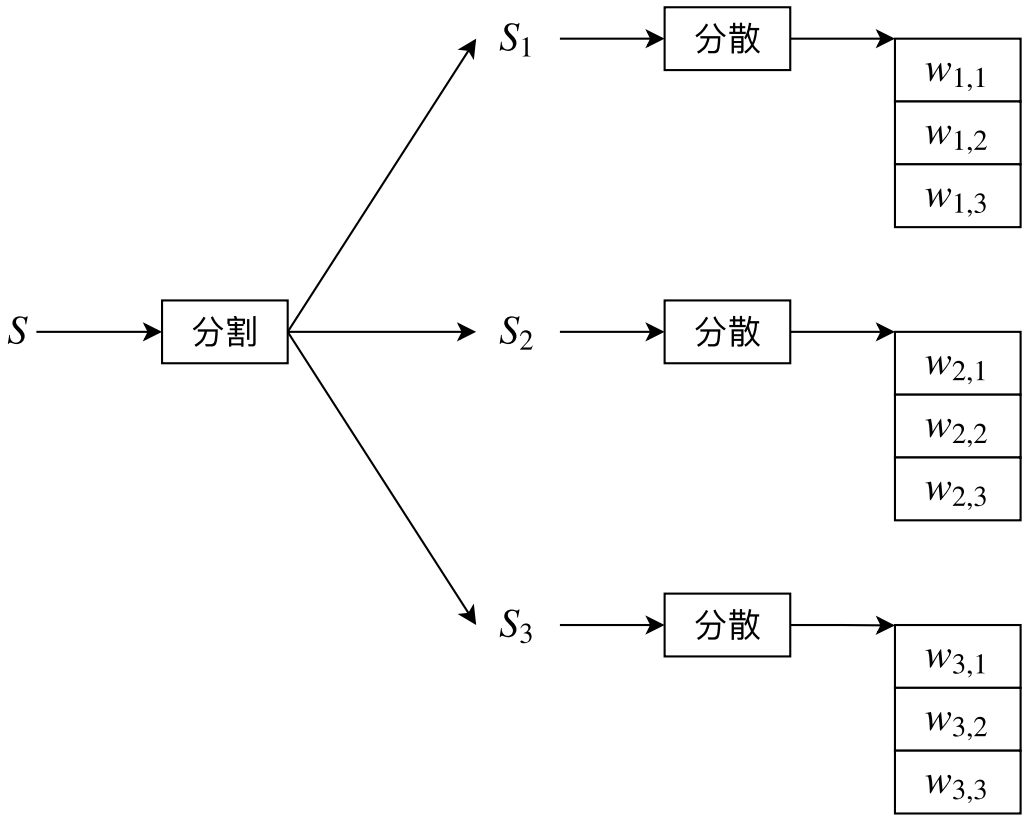

図 3.1　部分復元可能な秘密分散法の分散段階のデータの流れ

と表すことができ，各分割したデータ長分だけ左シフトして直和をとることで S となる．分割データ S_i の集合を

$$S_{all} = \{S_1, S_2, \cdots, S_d\}$$

とおき，復元したいデータの集合 S_c を S_{all} の部分集合，すなわち

$$S_{all} \supseteqq S_c = \{S_{c_1}, S_{c_2}, \cdots, S_{c_j}\}(1 \leqq j \leqq d)$$

とすると，復元したいデータ S' は

$$S' = S_{c_1} \prod_{m=2}^{j} 2^{l_m} + S_{c_2} \prod_{m=3}^{j} 2^{l_m} + \cdots + S_{c_j}$$

と表すことができる．S_c の任意の要素である $S_{c_t}(1 \leqq t \leqq j)$ のシェア集合を

$$W_{c_t} = \{w_{c_{t,1}}, w_{c_{t,2}}, \cdots, w_{c_{t,n}}\}$$

とおく．復元したいデータのシェア集合 W_{c_t} を集めたものをシェア集合 G_c

$$G_c = \{W_{c_1}, W_{c_2}, \cdots, W_{c_j}\}$$

とおき，シェア集合 G_c の要素 W_{c_t} を復元したい項目のみを含んだシェアとなるように結合するための紐付け情報

$$\mathbf{U}_c = \begin{pmatrix} \prod_{m=2}^{j} 2^{l_m} \\ \prod_{m=3}^{j} 2^{l_m} \\ \vdots \\ 1 \end{pmatrix} \tag{3.2}$$

を作成する．

S を式 (3.1) を用いて分割し，$S_i (i = 1, 2, \cdots, e)$ を作成する．S_i のデータサイズを l_i とする．素数 $p (S < p$ かつ $n < p)$ を選択する．$\mathbb{Z}/p\mathbb{Z} - \{0\}$ の集合 $X = \{x_1, x_2, \cdots, x_n\} (a \neq b$ のとき $x_a \neq x_b)$ から

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{k-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{k-1} \end{pmatrix} \pmod{p}$$

となるような $n \times k$ の vandermonde 行列 $\mathbf{X}$ を作成する．また分割データ S_i と $\mathbb{Z}/p\mathbb{Z}$ の集合 $R_i = \{r_{i,1}, r_{i,2}, \cdots, r_{i,k-1}\}$(ただし $r_{i,k-1} \in \mathbb{Z}/p\mathbb{Z} - \{0\}$) からランダムに選択し，ベクトル $\mathbf{A}_i$

$$\mathbf{A}_i = \begin{pmatrix} S_i \\ r_{i,1} \\ r_{i,2} \\ \vdots \\ r_{i,k-1} \end{pmatrix} (r_1, r_2, \cdots, r_{k-1} \in R, r_{k-1} \neq 0)$$

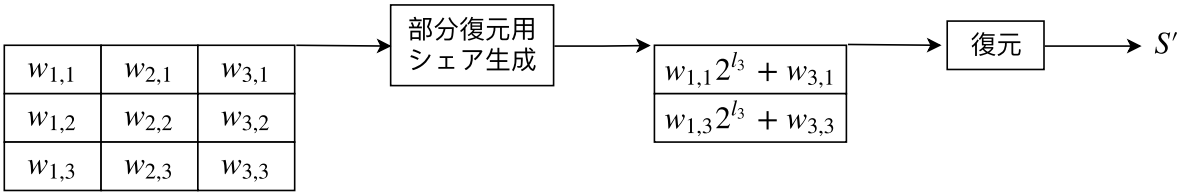

図 3.2　部分復元可能な秘密分散法の復元段階のデータの流れ

を作成する．p を法とした $\mathbf{X}$ と $\mathbf{A}_i$ の乗算より，

$$\mathbf{X}\mathbf{A}_i = \begin{pmatrix} w_{i,1} \\ w_{i,2} \\ \vdots \\ w_{i,n} \end{pmatrix} \pmod{p}$$

となるような $w_{i,j}(j = 1, 2, \cdots, n)$ をシェアとよぶ．$w_{i,j}$と x_i を対応付けて分散する．

3.2　復元段階

復元段階のデータの流れを図 3.2 に示す．G_c から各 w_{c_j} のシェアを k 個以上集める．集めたシェアから k 個選択し，シェア行列 $\mathbf{W}_c$

$$\mathbf{W}_c = \begin{pmatrix} w_{c_{1,1}} & w_{c_{2,1}} & \cdots & w_{c_{j,1}} \\ w_{c_{1,2}} & w_{c_{2,2}} & \cdots & w_{c_{j,2}} \\ \vdots & \vdots & \ddots & \vdots \\ w_{c_{1,k}} & w_{c_{2,k}} & \cdots & w_{c_{j,k}} \end{pmatrix}$$

を作成する．選択したシェアに対応する x_i より，

$$\mathbf{X}' = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{k-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_k & x_k^2 & \cdots & x_k^{k-1} \end{pmatrix} \pmod{p}$$

となるような $k \times k$ の vandermonde 行列 X' を作成する．$\mathbf{W}_c$ に対して式 (3.2) をかけ，部分復元用シェア $\mathbf{W}_c\mathbf{U}_c$ を作成する．

$$\mathbf{W}_c\mathbf{U}_c = \begin{pmatrix} w_{c_{1,1}} \prod_{m=2}^{j} 2^{l_m} + w_{c_{2,1}} \prod_{m=3}^{j} 2^{l_m} + \cdots + w_{c_{j,1}} \\ w_{c_{1,2}} \prod_{m=2}^{j} 2^{l_m} + w_{c_{2,2}} \prod_{m=3}^{j} 2^{l_m} + \cdots + w_{c_{j,2}} \\ \vdots \\ w_{c_{1,k}} \prod_{m=2}^{j} 2^{l_m} + w_{c_{2,k}} \prod_{m=3}^{j} 2^{l_m} + \cdots + w_{c_{j,k}} \end{pmatrix} \tag{3.3}$$

式 (3.3) を展開してまとめると，

$$\mathbf{X}' \begin{pmatrix} S_{c_1} \prod_{m=2}^{j} 2^{l_m} + S_{c_2} \prod_{m=3}^{j} 2^{l_m} + \cdots + S_{c_j} \\ r_{1,1} \prod_{m=2}^{j} 2^{l_m} + r_{2,1} \prod_{m=3}^{j} 2^{l_m} + \cdots + r_{j,1} \\ \vdots \\ r_{1,k-1} \prod_{m=2}^{j} 2^{l_m} + r_{2,k-1} \prod_{m=3}^{j} 2^{l_m} + \cdots + r_{j,k-1} \end{pmatrix} \tag{3.4}$$

となる．式 (3.4) に $\mathbf{X}'$ の逆行列 $\mathbf{X'}^{-1}$ を左からかけると

$$\mathbf{X'}^{-1}\mathbf{X}' \begin{pmatrix} S_{c_1} \prod_{m=2}^{j} 2^{l_m} + S_{c_2} \prod_{m=3}^{j} 2^{l_m} + \cdots + S_{c_j} \\ r_{1,1} \prod_{m=2}^{j} 2^{l_m} + r_{2,1} \prod_{m=3}^{j} 2^{l_m} + \cdots + r_{j,1} \\ \vdots \\ r_{1,k-1} \prod_{m=2}^{j} 2^{l_m} + r_{2,k-1} \prod_{m=3}^{j} 2^{l_m} + \cdots + r_{j,k-1} \end{pmatrix}$$

となり，

$$S' = S_{c_1} \prod_{m=2}^{j} 2^{l_m} + S_{c_2} \prod_{m=3}^{j} 2^{l_m} + \cdots + S_{c_j}$$

を復元できる．以上の手順より，分散したデータ S から一部のデータだけを復元する部分復元を実現している．

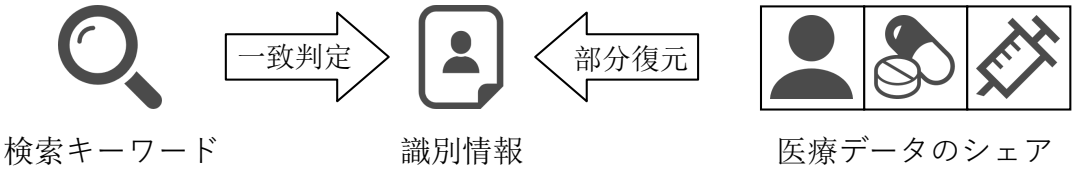

図 3.3　部分復元可能な秘密分散法の検索の流れ

3.3　部分復元可能な秘密分散法における検索

医師は診療を行う際，患者の医療データをバックアップデータから検索して閲覧する．しかし，部分復元可能な秘密分散法は検索の仕組みが無く，図 3.3 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復元してから検索キーワードと一致判定を行い検索する方法が考えられる．識別情報を部分復元するには医療データのシェアの組み合わせを知っている必要があるため，各ストレージで対応を持たせておく必要がある．しかし，医療データのシェアには x_i が対応づけて保存されており，医療データのシェアの組み合わせが分かれば医療データを不正に復元される恐れがある．また，復元は方程式を解くために逆行列を作成する必要があり，逆行列の作成には $\mathcal{O}(l^3)$(医療データのデータ長を l とする) の計算量がかかる．バックアップしている全員分の識別情報を復元してから検索を行うため時間を要する．そのため，多数の患者を素速く診療する必要がある災害時にはこの検索方法は適していない．したがって，検索にかかる計算量を削減し検索を高速化する必要がある．

3.4　まとめ

本章では，(k, n) しきい値秘密分散バックアップした医療データを災害時に活用するために提案された部分復元可能な秘密分散法について分散段階と復元段階に分けて述べ，部分復元可能な秘密分散バックアップした医療データの検索方法について述べた．部分復元可能な秘密分散法は分散した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復元してから検索する方法が考えられるが，復元するためには同一人物の医療データのシェアを対応づけて保存しておく必要があるため，医療データを不正に復元される恐れが

あり安全性での問題がある．また，復元は方程式を解くためにかかる計算量が大きく，全員分の識別情報を復元してから検索を行うため検索に時間を要する．そのため，多数の患者を素速く診療する必要がある災害時にはこの検索方法は適していない．したがって，検索にかかる計算量を削減し検索を高速化する必要がある．

第 4 章

シェア間の係数の差の比較による検索

安全性を落とすことなく部分復元可能な秘密分散法における検索の計算量の削減を目的として，患者の識別情報を復元すること無くシェアの状態で検索を可能とする方法を提案する．提案方法の検索の流れを図 4.1 に示す．提案方法は，識別情報を $(2, n)$ しきい値秘密分散して作成したタグを医療データのシェアに付与する．そして，検索キーワードをタグと同様に $(2, n)$ しきい値秘密分散して検索タグを作成し，検索タグと医療データのシェアに付与されているタグの係数の差を比較することでシェアの状態で検索を可能としている．

本章では，シェアの状態で検索を可能とする方法について述べ，医療データのシェアとタグの保存方法について述べた後，提案方法の評価を行う．

4.1 基本的な考え

部分復元可能な秘密分散法では検索の仕組みは無く，識別情報を部分復元してから検索を行う方法が考えられる．復元には方程式を解くため逆行列を作成する必要があり，$\mathcal{O}(l^3)$ の

図 4.1　シェア間の係数の差の比較による検索の流れ

計算量がかかり時間を要する．また識別情報を復元するためには同一人物の医療データのシェアの組み合わせを知っている必要があり，各ストレージで対応を持たせて保存しておく必要がある．しかし医療データのシェアの組み合わせが分かると，医療データを不正に復元される恐れがある．また，検索可能な秘密分散法の研究がなされているがタグを確定的に生成するため識別情報が一意でない医療データに適用すると頻度分析の耐性が無いことや，復元処理を行うため検索にかかる計算量が大きいなどの課題が挙げられる [10][11]．したがって，安全性を落とすことなく検索にかかる計算量を削減する必要がある．復元処理に $\mathcal{O}(l^3)$ の計算量がかかっていることから，識別情報を復元すること無くシェアの状態で検索が可能であれば高速化が見込める．また，医療データのバックアップは秘密分散法を用いているため，検索の仕組みに暗号などの方式を用いると安全性が暗号に依存する可能性がある．そこで，本研究では秘密分散法をベースに検索の仕組みを考える．

4.2　分散段階

分散段階のデータの流れを図 4.2 に示す．分散バックアップしたい医療データに対応する識別情報を kw $(0 < kw)$ とする．素数を p $(kw < p, n < p)$ とし，$\mathbb{Z}/p\mathbb{Z} - \{0\}$ 上の集合を X とする．また，p, X は医療データを秘密分散した際とは異なる値にする．X は

$$X = \{x_1, x_2, \cdots, x_n\} (i \neq j \text{ のとき } x_i \neq x_j)$$

とすることができ，X より

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \pmod{p}$$

となるような vandermonde 行列 $\mathbf{X}$ を作成する．kw と $\mathbb{Z}/p\mathbb{Z} - \{0\}$ 上の乱数 r より

$$\mathbf{A} = \begin{pmatrix} kw \\ r \end{pmatrix}$$

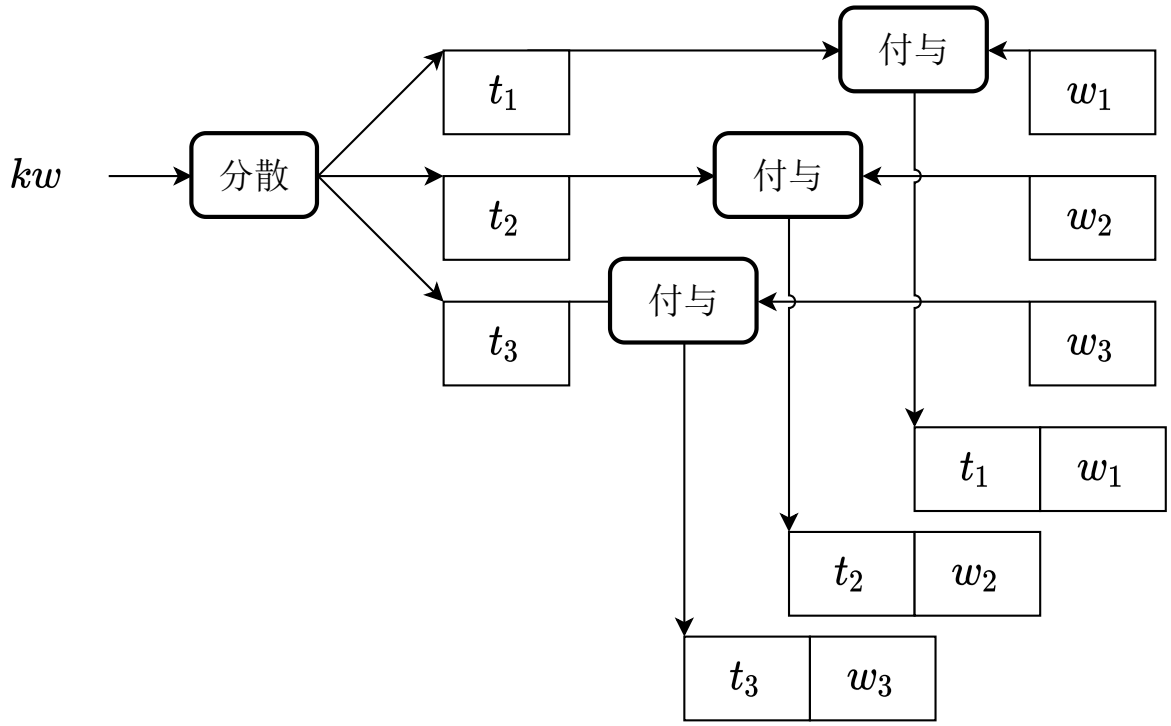

図 4.2　シェア間の係数の差の比較による検索の分散段階のデータの流れ

となるようなベクトル $\mathbf{A}$ を作成する．p を法とする $\mathbf{X}$ と $\mathbf{A}$ の乗算より

$$\mathbf{XA} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{pmatrix} \pmod{p}$$

となるようなタグ $t_u(u = 1, 2, \cdots, n)$ が得られる．また，方程式で表すと

$$t_u \equiv kw + rx_u \pmod{p} \tag{4.1}$$

となる．タグと医療データのシェアを対応づけて分散バックアップする．同時に，集合 X と p を保持しておく．

4.3　検索段階

分散段階のデータの流れを図 4.3 に示す．キーワード kw' を検索する場合を考える．

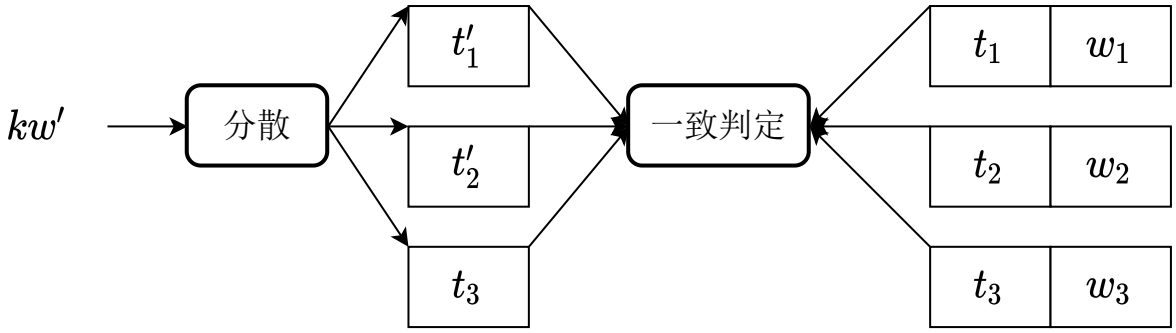

図 4.3　シェア間の係数の差の比較による検索の復元段階のデータの流れ

$\mathbb{Z}/p\mathbb{Z} - \{0\}$ 上の乱数 r' と kw' より

$$\mathbf{A}' = \begin{pmatrix} kw' \\ r' \end{pmatrix}$$

となるようなベクトル $\mathbf{A}'$ を作成する．分散段階で用いた行列 $\mathbf{X}$ と $\mathbf{A}'$ の p を法とした乗算より

$$\mathbf{X}\mathbf{A}' = \begin{pmatrix} t'_1 \\ t'_2 \\ \vdots \\ t'_n \end{pmatrix} \pmod{p}$$

となるような検索タグ t'_u が得られる．また，方程式で表すと

$$t'_u \equiv kw' + r'x_u \pmod{p} \tag{4.2}$$

となる．

ここで，医療データのシェアに付与されているタグ t_u と検索タグ t'_u の減算を行う．式 (4.1)(4.2) より

$$t_u - t'_u \equiv (kw - kw') + (r - r')x_u \pmod{p} \tag{4.3}$$

となる．さらに，式 (4.3) の両辺に x_u の逆元 x_u^{-1} を掛けると

$$(t_u - t'_u)x_u^{-1} \equiv (kw - kw')x_u^{-1} + (r - r') \pmod{p} \tag{4.4}$$

となる．$kw = kw'$ の場合，式 (4.4) より係数の差 $r - r'$ 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，係数の差 $r - r'$ を正しく求め

ることができず，全ての u でこの結果が同じ値になることは無い．これを利用してシェアの状態で一致判定を行い，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 $\mathcal{O}(l^3)$ 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．

4.3.1 一致判定が可能であることの証明

提案方法は，タグと検索タグの係数の差を求めることで一致判定を行っている．そこで，本項では，$kw = kw'$ の場合式 (4.4) は全ての u で同じ値になり，$kw \neq kw'$ の場合式 (4.4) は全ての u で異なる値になることの証明を行う．

$kw = kw'$ の場合，

$$(kw - kw')x_u^{-1} = 0$$

であるため，式 (4.4) より係数の差 $r - r'$ は正しく求まる．また，全ての u でこの結果は同じ値になる．

一方 $kw \neq kw'$ の場合，n 個のうち 2 個 t_i，t_j と t'_i，t'_j $(i, j \in u)$ を考える．式 (4.4) より，

$$(t_i - t'_i)x_i^{-1} \equiv (kw - kw')x_i^{-1} + (r - r') \pmod{p} \tag{4.5}$$

$$(t_j - t'_j)x_j^{-1} \equiv (kw - kw')x_j^{-1} + (r - r') \pmod{p} \tag{4.6}$$

となる．このとき，$(t_i - t'_i)x_i^{-1} \equiv (t_j - t'_j)x_j^{-1} \pmod{p}$ であると仮定すると式 (4.5，4.6) より，

$$(kw - kw')x_i^{-1} + (r - r') \equiv (kw - kw')x_2^{-1} + (r - r') \pmod{p} \tag{4.7}$$

となる．式 (4.7) を整理すると，

$$x_i^{-1} \equiv x_j^{-1} \pmod{p}$$

となる．$x_i, x_j \in \mathbb{Z}/p\mathbb{Z} - \{0\}$ であるため，$x_i = x_j$ となる．これは集合 X の条件 ($i \neq j$ のとき $x_i \neq x_j$) に反する．したがって，すべての u についても同様にいうことができ $kw \neq kw'$ の場合式 (4.4) の結果が同じ値になることはない．また，以上より n 個すべてのタグと検索タグを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．

表 4.1　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
⋮	⋮	⋮

4.3.2　医療データのシェアとタグの保存方法

提案方法ではタグと検索タグの係数の差を求めるために，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り，各ストレージで対応を持たせて保存しておく必要がある．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かってしまうため，医療データを不正復元される恐れがある．そのため，同一の識別情報から作成した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は分からないような仕組みが必要となる．

タグと検索タグの比較は n 個行う必要はなく 2 個行えば良いことから，表 4.1 に示すように 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．また，各ストレージで医療データのシェアに同一のタグが付与されていると対応が分かってしまう可能性がある．そのため各ストレージで医療データのシェアに付与するタグは異なるタグにする必要があり，医療データのしきい値が (k, n) とするとタグのしきい値を $(2, 2n)$ とすることで異なるタグを付与することができ，添字が u のストレージは表 4.2 のようになる．この保存方法により，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，一致判定をするために必要なタグ 2 個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は 2 個以上知ることはできない．したがって，第三者にストレージにアクセスされたとしても医療データを復元することはできない．

表 4.2　添字が u のストレージ

医療データのシェア	タグ $_{2u-1}$	タグ $_{2u}$
A さんシェア $_u$	A さん t_{2u-1}	A さん t_{2u}
B さんシェア $_u$	B さん t_{2u-1}	B さん t_{2u}
⋮	⋮	⋮

4.4　まとめ

部分復元可能な秘密分散法における検索の高速化を目的として，患者の識別情報を復元すること無くシェアの状態で検索を可能とする方法を提案した．提案方法は，識別情報を $(2, n)$ しきい値秘密分散して作成したタグを医療データのシェアに付与する．そして，検索キーワードをタグと同様に $(2, n)$ しきい値秘密分散して検索タグを作成し，検索タグと医療データのシェアに付与されているタグの係数の差を比較することでシェアの状態で検索を可能としている．

本章では，シェアの状態で検索を可能とする方法について述べ，医療データのシェアとタグの保存方法について述べた後，提案方法の評価を行った．提案方法により，従来法では一致判定に $\mathcal{O}(l^3)$ の計算量がかかっていたところ，提案法では $\mathcal{O}(d^2)$ の計算量となる．また，提案方法はタグと検索タグの添字が同じもの同士を正しく比較する必要があり，同一の識別情報から作成したタグの組み合わせが分かっている必要がある．タグは医療データのシェアに対応づけて保存されており医療データのシェアの組み合わせが分かってしまうため，第三者がストレージにアクセスされた場合医療データを不正に復元される恐れがある．そこで，提案したシェアとタグの保存方法を用いることで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り，一致判定をするために必要なタグ 2 個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は 2 個以上知ることはできない．したがって，第三者にストレージにアクセスされたとしても医療データを復元することはできない．

第 5 章

評価

安全性を落とすことなく部分復元可能な秘密分散法の検索にかかる計算量の削減を目的として，シェア間の係数の差の比較による検索を提案した．識別情報を復元する従来の方法では $\mathcal{O}(l^3)$ の計算量がかかっていたところ，提案方法では $\mathcal{O}(d^2)$ の計算量に削減した．しかし，ある条件下においてタグの作成に使用した識別情報 kw が有限個に絞られる．

本章では，識別情報を復元する従来の方法と提案方法の検索にかかる時間を比較する．その後，識別情報 kw が有限個に絞られる条件について述べ，提案方法の安全性について評価する．

5.1 従来法との検索時間の比較

復元処理を不要にしたことによる検索時間の変化を調べる．医師によると 15 秒程度で医療データが手元に欲しいということから，高知県の人口である 70 万人分のデータを 15 秒以内に検索することを目標とする．データ数は 1 人につき 1 個データがあるとし，10 万個から 70 万個まで計測している．識別情報と検索キーワードは名前を想定して 144bit(9 文字の平仮名) とし，全体のデータは医療データを想定して 3000byte とした．分散段階では，データのシェアは $(2,3)$ しきい値秘密分散法で作成し，タグは $(2,6)$ しきい値秘密分散法で作成しデータのシェアに付与した．分散段階の処理の流れを図 5.1 に示す．従来の方法では識別情報を部分復元し検索キーワードと一致判定を行い結果を出力するのにかかった時間を計測している．従来の方法の検索処理の流れを図 5.2 に示す．提案方法では検索キーワードを $(2,6)$ しきい値秘密分散してタグを作成し，タグと検索タグの一致判定を行い結果を出力

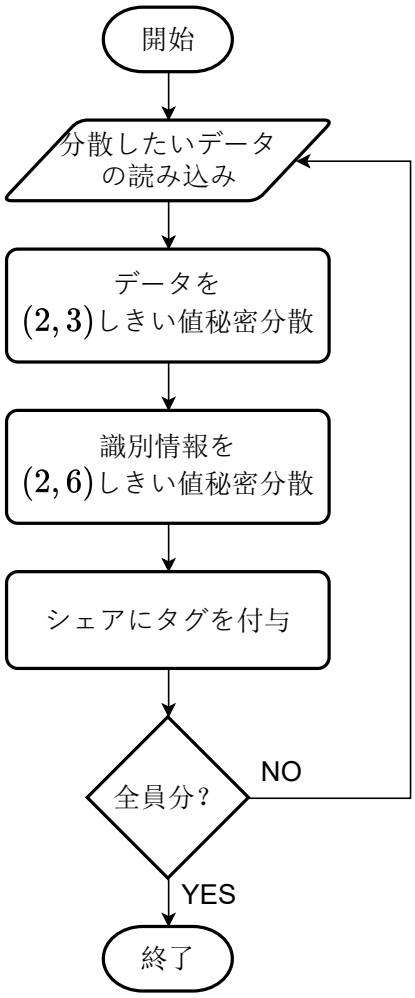

図 5.1　分散段階の処理の流れ

するのにかかった時間を計測している．提案方法の検索処理の流れを図 5.3 に示す．時間の計測は，プログラムの最初と最後に std::chrono 関数を用いてシステム時間を取得し，その差分を計算した．実験環境を表 5.1 に示す．

5.1.1　結果

従来法と提案法の検索にかかった時間を図 5.4 に示す．同じ処理をデータ数分繰り返しているためグラフはデータ数に比例して線形になっており，目標である 70 万人分のデータを 15 秒以内に検索できていることが確認できる．また，同じ検索時間内で検索できるデータ数の差は検索時間に比例して広がっていくと考えられる．

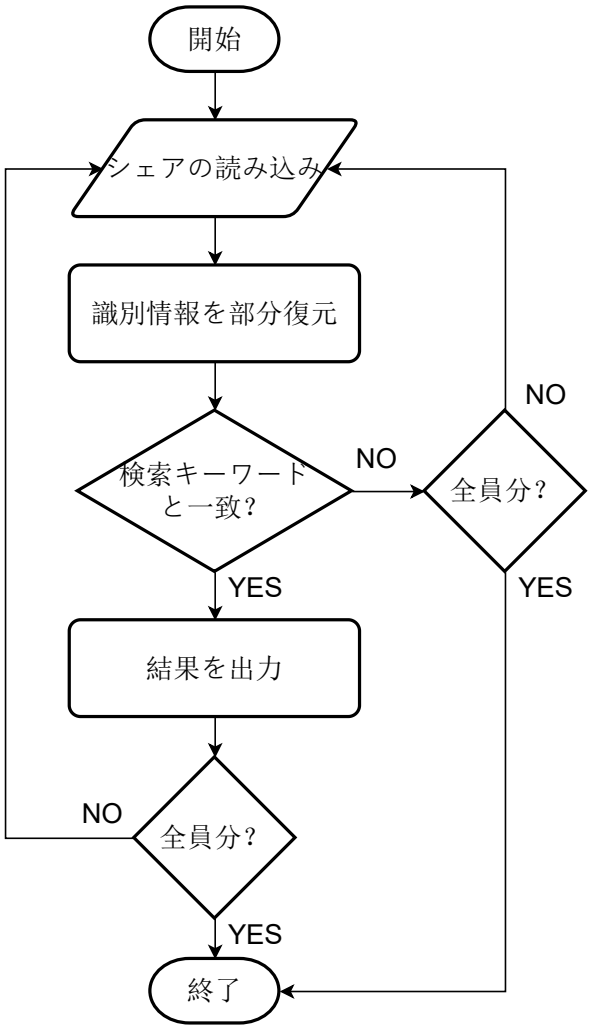

図 5.2　従来法の検索処理の流れ

5.1.2　考察

図 5.4 より，70 万個のデータを 15 秒以内に検索するという目標を達成できていることが確認できた．しかし，今回用いたデータはダミーデータであり実際に医療データが何個になるかは定かではないため，さらなる高速化が必要になる可能性がある．高速化の手法としては，有限体上で演算を行っているため拡大体に拡張することや，並列化を行うことでさらなる高速化が見込めると考える．

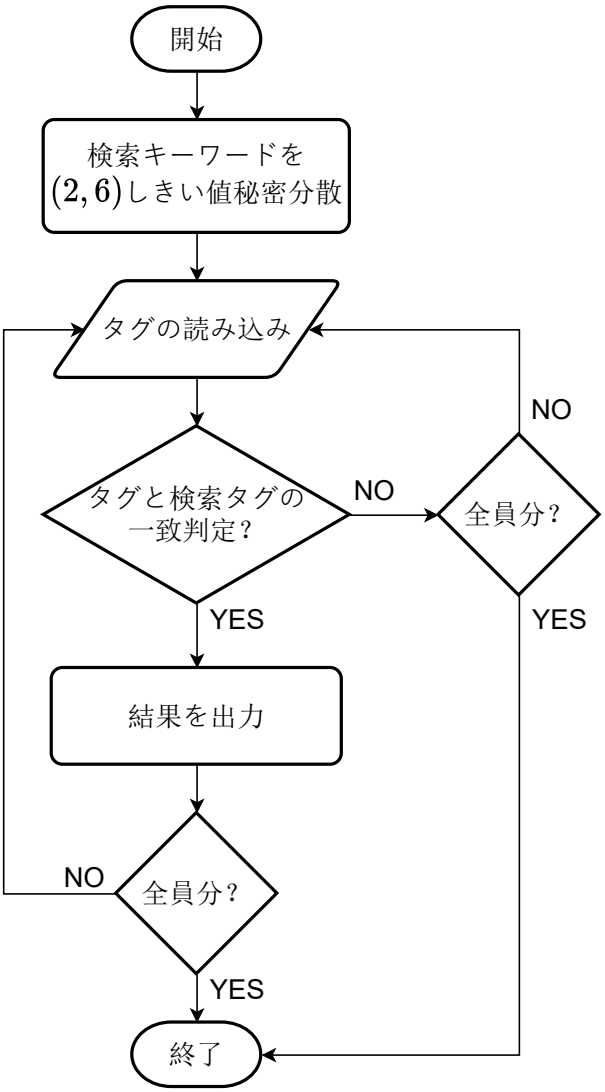

図 5.3　提案法の検索処理の流れ

5.2　安全性

提案方法では，タグは医療データのシェアに対応付けて 2 個ずつストレージに保存し，各ストレージでランダムに保存されているため，攻撃者がストレージにアクセス可能であっても医療データのシェアの組み合わせは分からず，タグからも組み合わせは分からないため医療データを不正に復元される恐れはない．しかし，アクセスしたシェアに対応するタグ t_u のうち 2 個手に入れることができる．また，提案方法はタグを作成する際のしきい値 k は 2 で固定であるため，攻撃者がこの検索方法を使っていると知っている場合には k は明らかで

表 5.1　実験環境

OS	Ubuntu 20.04.1LTS
CPU	Intel ®Xeon W-1290P 3.7GHz
メモリ	16GB
コンパイラ	g++-9.3.0
コンパイラオプション	-O，-std=c++11

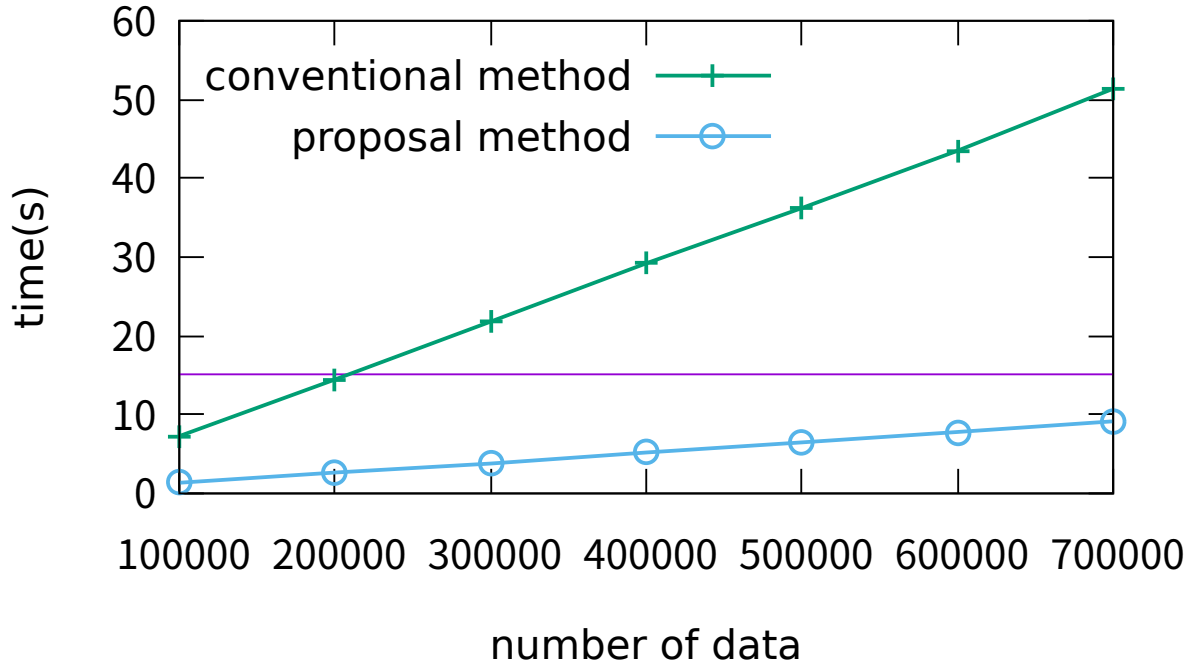

図 5.4　従来法と提案法の検索時間の比較

ある．一方集合 X，素数 p は全てのタグで共通であるためタグに対応付けて保存する必要はなく，医師が使用する端末に保存しておくことが考えられる．しかし，災害時では仮設の診療所でタブレット端末などを用いることが想定されているため，端末の盗難などにより p, X が攻撃者に漏洩する可能性がある．

本節では，以下に示す条件について識別情報 kw が有限個に絞られる方法について検討する．

1. t_1，t_2，p，k が漏洩
2. t_1，t_2，x_1，x_2，k が漏洩

5.2.1 条件 1：t_1，t_2，p，k が漏洩

t_1，t_2 から kw を求めるためには p，k に加えて x_1，x_2 があれば求めることができる．t_1，t_2 の作成の条件より $x_1, x_2 \in \mathbb{Z}/p\mathbb{Z} - \{0\}$，$x_1 \neq x_2$ であるため，x_1，x_2 の全てのパターン $(p-1)(p-2)$ 通りを調べることで kw を有限個に絞り込まれる．以下に kw を絞り込む手順を示す．t_1, t_2 より，

$$\mathbf{T} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$$

となるようなベクトル $\mathbf{T}$ を生成する．x_1, x_2 は変数として扱い，

$$\mathbf{X}' = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \end{pmatrix}$$

となるような行列 $\mathbf{X}'$ を生成し，逆行列 $\mathbf{X}'^{-1}$ を求める．

$$\mathbf{X}'^{-1} = \frac{1}{x_2 - x_1} \begin{pmatrix} x_2 & -x_1 \\ -1 & 1 \end{pmatrix}$$

$\mathbf{X}'^{-1}$ を $\mathbf{T}$ の左側から掛けると

$$\begin{aligned} \mathbf{X}'^{-1}\mathbf{T} &= \frac{1}{x_2 - x_1} \begin{pmatrix} x_2 & -x_1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \\ &= \begin{pmatrix} \dfrac{t_1 x_2 - t_2 x_1}{x_2 - x_1} \\ \dfrac{t_2 - t_1}{x_2 - x_1} \end{pmatrix} \end{aligned}$$

となる．kw は，

$$kw \equiv \frac{t_1 x_2 - t_2 x_1}{x_2 - x_1} \pmod{p} \tag{5.1}$$

である．式 (5.1) に対して x_1, x_2 の全てのパターンを試すことで kw を有限個に絞り込まれる．

以下に例を示す．正解の組み合わせを式 (5.2) のようにし，

$$\begin{cases} 0 \equiv 2 + 3 \times 1 \pmod{5} \\ 1 \equiv 2 + 3 \times 3 \pmod{5} \end{cases} \tag{5.2}$$

ここから $t_1 = 0$，$t_2 = 1$，$p = 5$，$k = 2$ が漏洩したとすると攻撃者は式 (5.3) のような連立合同式が得られる．

$$\begin{cases} 0 \equiv kw + r \times x_1 \pmod 5 \\ 1 \equiv kw + r \times x_2 \pmod 5 \end{cases} \tag{5.3}$$

x_1, x_2 を変数として扱い，式 (5.3) を kw について解くと

$$kw \equiv \frac{0 \times x_2 - 1 \times x_1}{x_2 - x_1} \pmod 5 \tag{5.4}$$

となる．式 (5.4) に対し，x_1, x_2 の組み合わせを 4×3 通り試すと，$kw = 2, 3, 4$ の 3 通りに絞られる．

5.2.2　考察

t_1，t_2，p，k が漏洩した場合，x_1，x_2 の全ての組み合わせ $(p-1)(p-2)$ 通りを調べることで kw が有限個に絞られる．t_1，t_2 を方程式で表すと

$$\begin{cases} t_1 \equiv kw + rx_1 \pmod p \\ t_2 \equiv kw + rx_2 \pmod p \end{cases} \tag{5.5}$$

となる．このとき

$$t_1 \equiv kw \pmod p$$

であるとすると，

$$rx_1 \equiv 0 \pmod p$$

となる．$r, x_1 \in \mathbb{Z}/p\mathbb{Z} - \{0\}$ であるため，満たす r, x_1 は存在しない．また，

$$t_2 \equiv kw \pmod p$$

についても同様のことが言える．したがって，式 (5.5) を解いた結果に対して x_1, x_2 の組み合わせを全て調べると kw は t_1, t_2 の値以外の $p-2$ 通り以下に絞られる．このことから p を大きくすることで絞られる範囲を広げることが可能である．また，識別情報を絞ってから医療データを復元するためには識別情報を一意に定め，各ストレージでランダムに保存されている医療データのシェアを正しい組み合わせで k 個以上集めてから復元する必要があるた

め，医療データを復元することは困難である．しかし，識別情報を絞られた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，十分に安全であるとは言えず p をシステムで秘匿し安全性を担保する必要がある．

5.2.3　条件 2：t_1，t_2，x_1，x_2，k が漏洩

t_1，t_2，x_1，x_2，k が漏洩しているため，連立合同式を解くことは可能であるが剰余演算の法 p が分からないため，kw を定めることができない．しかし，除算を行わず連立合同式を解いた結果に対し p を総当りするとある条件下において kw を絞られる．以下に kw を絞り込む手順を示す．t_1，t_2 より，

$$\mathbf{T} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$$

となるような行列 $\mathbf{T}$ を生成する．x_1, x_2 より，

$$\mathbf{X}' = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \end{pmatrix}$$

となるような行列 $\mathbf{X}'$ を生成し，除算を行わず逆行列 $\mathbf{X}'^{-1}$ を求める．

$$\mathbf{X}'^{-1} = \frac{1}{x_2 - x_1} \begin{pmatrix} x_2 & -x_1 \\ -1 & 1 \end{pmatrix}$$

除算を行わず $\mathbf{X}'^{-1}$ を $\mathbf{T}$ の左側から掛けると

$$\begin{aligned} \mathbf{X}'^{-1}\mathbf{T} &= \frac{1}{x_2 - x_1} \begin{pmatrix} x_2 & -x_1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \\ &= \begin{pmatrix} \dfrac{t_1 x_2 - t_2 x_1}{x_2 - x_1} \\ \dfrac{t_2 - t_1}{x_2 - x_1} \end{pmatrix} \end{aligned}$$

となる．kw は，

$$kw \equiv \frac{t_1 x_2 - t_2 x_1}{x_2 - x_1} \pmod{p} \tag{5.6}$$

である．このとき，$t_1, t_2 < p$ なる p を変化させ kw を推測する．$\dfrac{t_1 x_2 - t_2 x_1}{x_2 - x_1}$ が自然数に割り切れる場合，つまり

$$kw \equiv a \pmod{p} \ (a \in \mathbb{N})$$

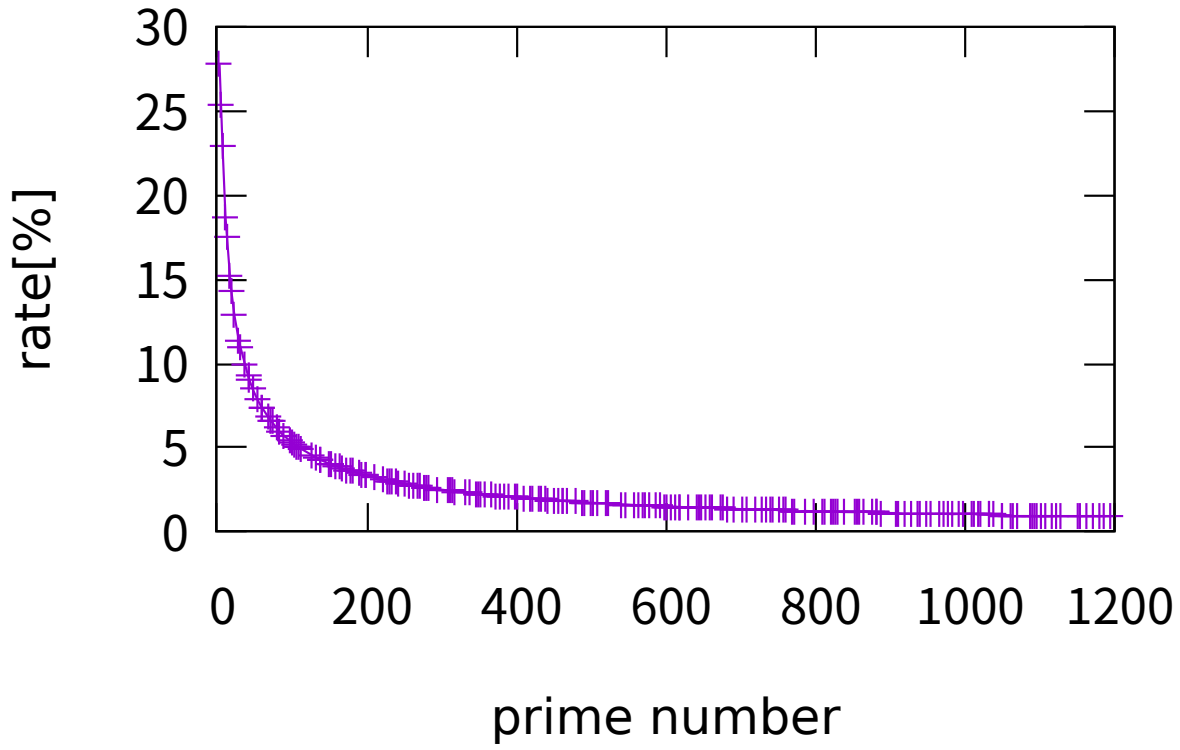

図 5.5　自然数に割り切れる確率

とすることが出来る場合，$a < p$ なる p では式 (5.6) より kw は全て a になる．したがって，kw は $t_1, t_2 < p < a$ なる素数 p の個数+1 通りに絞られる．$\dfrac{t_1x_2 - t_2x_1}{x_2 - x_1}$ が自然数に割り切れない場合は kw は不定である．

5.2.4　考察

式 (5.6) における $\dfrac{t_1x_2 - t_2x_1}{x_2 - x_1}$ が取る最大値は $(p-1)^2$ であり，kw は $t_1, t_2 < p < (p-1)^2$ なる素数 p の個数+1 通りに絞られる．このことから p を大きくすることで絞られる範囲を広げることが可能であり，図 5.5 に示すように $\dfrac{t_1x_2 - t_2x_1}{x_2 - x_1}$ が自然数に割り切れる確率も無視できるほど小さくなると考えられる．また，識別情報を絞ってから医療データを復元するためには識別情報を一意に定め，各ストレージでランダムに保存されている医療データのシェアを正しい組み合わせで k 個以上集めてから復元する必要があるため，医療データを復

元することは困難である．しかし，識別情報を絞られた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，十分に安全であるとは言えず t_1，t_2，x_1，x_2 をシステムで秘匿し安全性を担保する必要がある．

5.3　まとめ

提案方法はある条件下において識別情報 kw が有限個に絞られる．

本章では識別情報を復元する従来の方法と提案方法の検索にかかる時間を比較し，識別情報 kw が有限個に絞られる条件について述べ，提案方法の安全性を評価した．検索にかかる時間は 70 万人分のデータを 15 秒以内に検索するという目標を達成することができた．しかし，今回用いたデータはダミーデータであり実際に医療データが何個になるかは定かではないため，さらなる高速化が必要になる可能性がある．高速化の手法としては，有限体上で演算を行っているため拡大体に拡張することや，並列化を行うことでさらなる高速化が見込めると考える．また，識別情報 kw が有限個に絞られる条件に対して p を大きくすることで絞られる範囲を広げることが可能である．さらに，識別情報を絞ってから医療データを復元するためには識別情報を一意に定め，各ストレージでランダムに保存されている医療データのシェアを正しい組み合わせで k 個以上集めてから復元する必要があるため，医療データを復元することは困難である．しかし，識別情報を有限個に絞り込ま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，p，X をシステムで秘匿することで安全性を担保する必要がある．

第 6 章

医療データ検索システム [12]

シェア間の係数の差の比較による検索では，タグ t_1，t_2，しきい値 k と素数 p あるいは x_1，x_2 が漏洩した場合，kw が有限個に絞られる．これに対し，素数 p を大きくすれば問題ないが識別情報を有限個に絞り込ま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，p，X をシステムで秘匿することで安全性を担保する必要がある．

本章では，t_u，p，X を秘匿する検索システムの構成について述べる．

6.1　システムの構成

検索システムはストレージ群，リファレンスモニタ，情報端末，情報閲覧端末によって構成される．ストレージ群には医療データのシェアやタグをバックアップ保管しておく．また，リファレンスモニタはストレージ群にアクセスして検索を行い，ヒットした医療データのシェアを復元し，情報端末へ送信する．情報端末は，検索要求をリファレンスモニタに送信し，検索にヒットした医療データを受信する．受信した医療データは情報閲覧端末より情報端末にリモートアクセスし閲覧することが出来る．以上の内容を踏まえたシステム構成を図 6.1 に示す．リファレンスモニタ，情報端末，情報閲覧端末を設置したシステムの検索手順を以下に示す．

まず，医師は情報閲覧端末を用いて情報端末へリモートアクセスする．次に，検索要求をリファレンスモニタに送信する．さらに，リファレンスモニタは送られた検索要求より，ストレージ群へアクセスし医療データを検索する．最後に，リファレンスモニタは検索してヒットした医療データを復元し，情報端末に送信する．以上の手順により，医師は患者の医

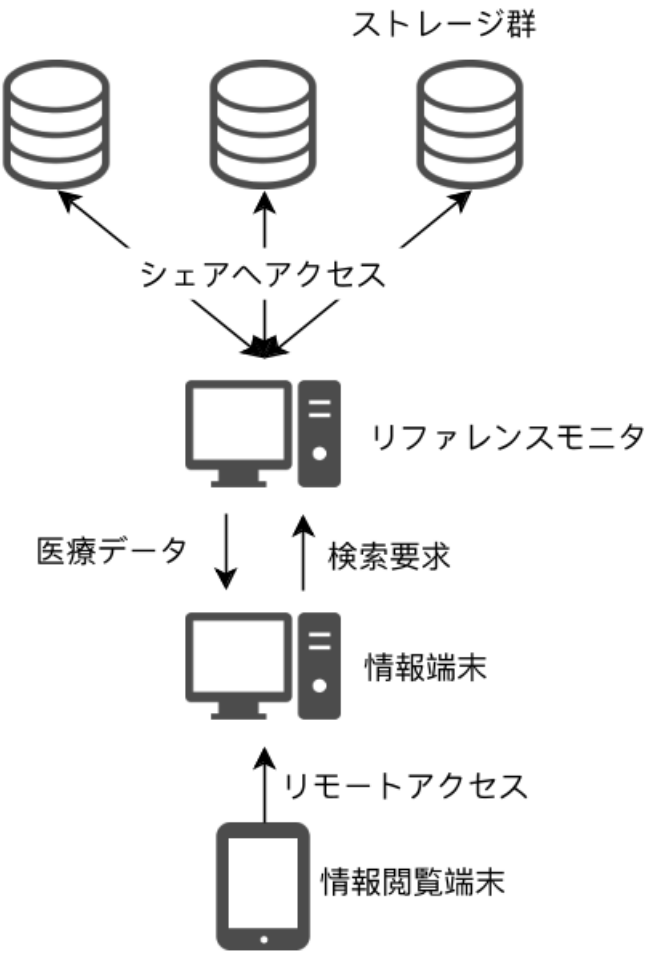

図 6.1　検索システムの構成

療データを閲覧することができる．

6.1.1　情報閲覧端末

医師が使用する情報閲覧端末は仮設の診療所でタブレット端末などを用いることが想定されているため，盗難や紛失の危険性や使いまわしをすることなどを考えると医療データを含むいかなる情報も保持しておくことは好ましくない．そのため，情報端末へリモートアクセスして医療データの検索・閲覧を行う．情報端末へリモートアクセスするためには利用者が正当であるかどうか，端末が正当であるかどうかを確認する必要がある．実際には，情報閲覧端末を配布する際に利用者が医師であることを医師免許等で確認した上で，パスワードを

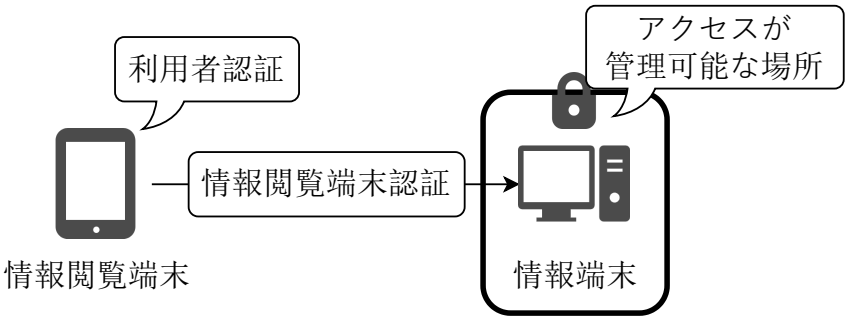

図 6.2　情報閲覧端末と情報端末間の具体的な対策

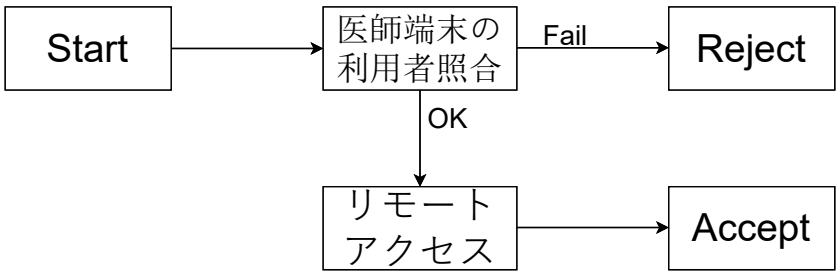

図 6.3　リモートアクセス時の状態遷移

設定するか指紋登録などを行い，情報端末へあらかじめ情報閲覧端末が正当であるかどうか確認するための認証情報を持たせておく．情報閲覧端末と情報端末間の具体的な対策を図 6.2 に示す．また，情報閲覧端末を用いて情報端末へリモートアクセスする際の状態遷移を図 6.3 に示す．

6.1.2　情報端末

第三者に自由な検索要求を許す場合，複数回の検索要求を行うことで不正に医療データを入手される可能性がある．そのため，検索要求を送信できる端末を情報端末に限定する．情報端末に復元済みの医療データを一時的に保持することで情報閲覧端末に医療データを保持することなくリモートアクセスで閲覧が可能となる．また，閲覧済みの医療データや検索履歴を不正に利用される恐れがあるため，医師が閲覧を終了した際 (リモートアクセスを終了した際) に閲覧済み医療データと検索履歴を削除する必要がある．情報端末からは情報閲覧端末の紛失や盗難があった場合，情報閲覧端末の利用者が正当であるかどうか確認すること

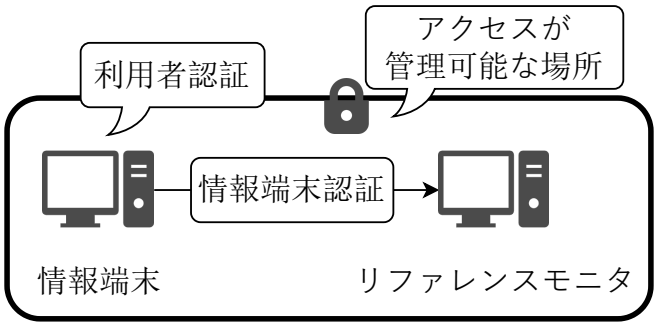

図 6.4　情報閲覧端末とリファレンスモニタ間の具体的な対策

は出来ないため，端末を使用する際には利用者認証を行う必要がある．また，リファレンスモニタへ検索要求を送信する際には，情報端末が正当であるかどうかを確認する必要がある．実際には，あらかじめ登録しておいたパスワードなどを用いて利用者の認証を行い，リファレンスモニタへあらかじめ情報端末が正当であるかどうかを確認するための認証情報を持たせておく．また，情報端末は拠点病院内のアクセスが管理可能な場所に設置することで，端末自体の安全性を確保する．情報端末とリファレンスモニタ間の具体的な対策を図 6.4 に示す．また，情報端末からリファレンスモニタへ検索要求を送信する際の状態遷移を図 6.5 に示す．

6.1.3　リファレンスモニタ

第三者にストレージへの自由なアクセスを許す場合，医療データを不正に復元される恐れがある．そのため，ストレージへのアクセスをリファレンスモニタに限定し，リファレンスモニタが医療データの検索・復元を行う．検索に必要なデータ X(タグ)，p(タグ) や医療データの復元に必要な p(医療データ) はリファレンスモニタが保持する．ストレージへアクセスする際には，リファレンスモニタが正当であるかどうかを確認する必要がある．実際には，ストレージへあらかじめリファレンスモニタが正当であるかどうかを確認するための認証情報を持たせておく．また，リファレンスモニタは拠点病院内のアクセスが管理可能な場所に設置することで，端末自体の安全性を確保する．リファレンスモニタとストレージ間の

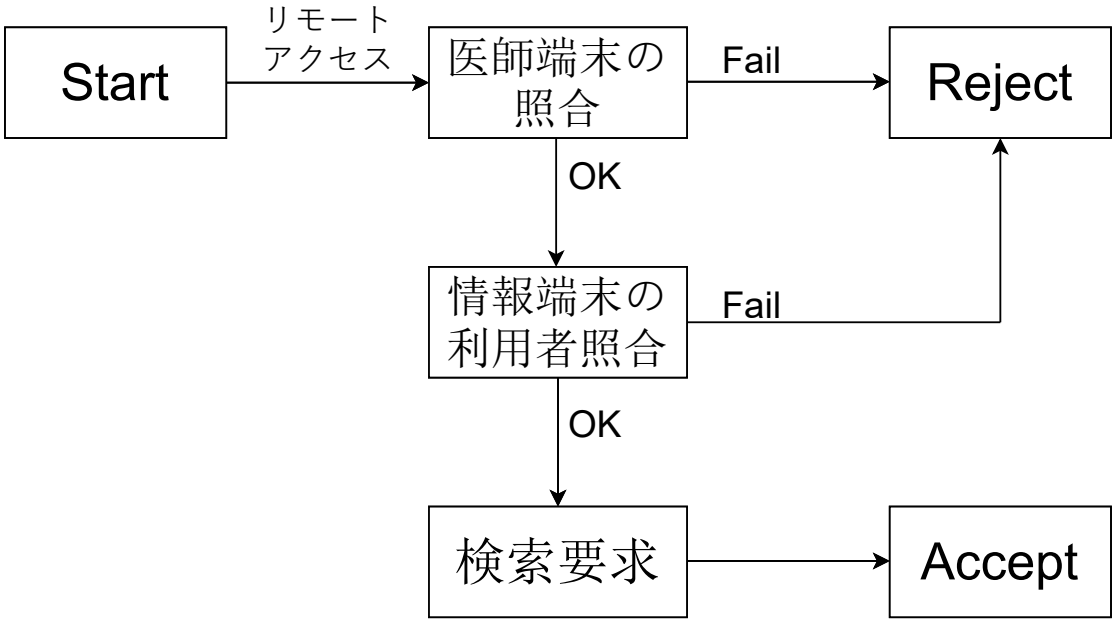

図 6.5　検索要求送信時の状態遷移

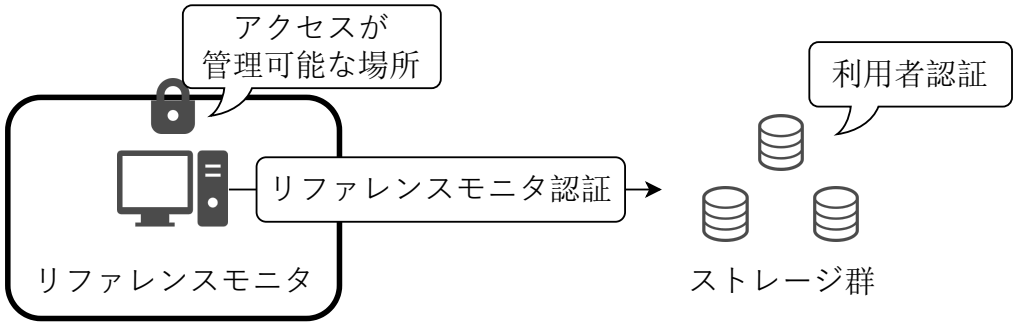

図 6.6　リファレンスモニタとストレージ間の具体的な対策

具体的な対策を図 6.6 に示す．また，リファレンスモニタからストレージへアクセスする際の状態遷移を図 6.7 に示す．

6.1.4　ストレージ群

ストレージ群には医療データのシェアやそれに対応するタグを保存する．ストレージには様々な病院の医療データが混在しているため，平時での利用も考えると利用者によってアクセスできる医療データを制御する必要がある．実際には，ストレージへあらかじめ利用者が

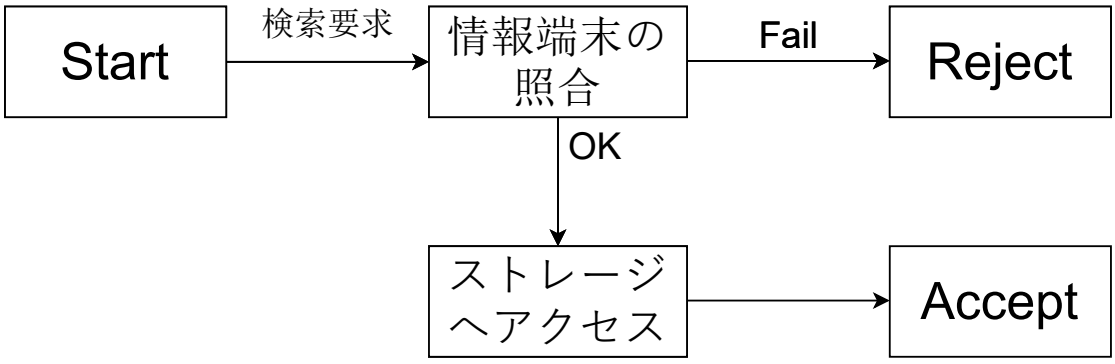

図 6.7　ストレージアクセス時の状態遷移

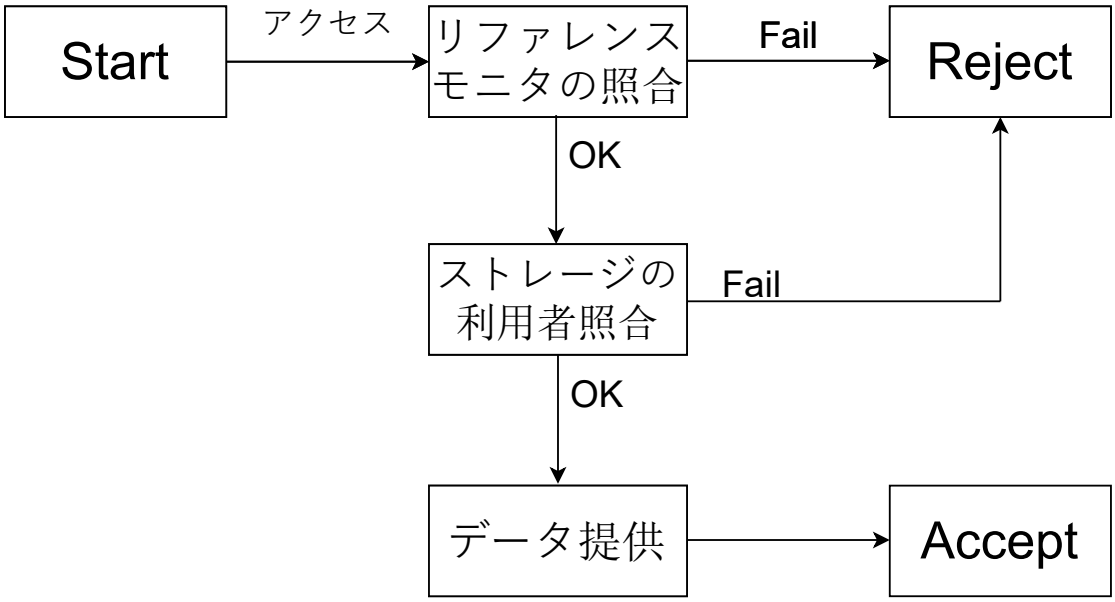

図 6.8　リファレンスモニタへ医療データ提供時の状態遷移

どの医療データにアクセス可能であるかを制御する認証情報を持たせておく．ストレージで医療データをリファレンスモニタへ提供する場合の状態遷移を図 6.8 に示す．

6.2　評価

提案システムでは，t_u，p，X を秘匿することで提案検索方法の安全性を高めている．提案システムに対して，リファレンスモニタのなりすましによるストレージへのアクセス，盗難などによる情報閲覧端末の操作，情報閲覧端末のなりすましが考えられる．

ストレージにアクセスすることが可能であっても，医療データのシェアやそれに対応するタグを入手することが可能であるが，提案検索方法を用いている場合シェアとタグの保存方法により医療データのシェアの組み合わせは分からず，タグからも医療データのシェアの組み合わせが分かることはないため，医療データを不正に復元される恐れはない．また，医師端末の操作やなりすましが可能であっても，情報閲覧端末にはリモートアクセス先の情報端末の情報が保存されているが，情報端末を利用するには利用者認証が必要なため検索要求などを送ることはできない．

これより，シェア間の係数の差の比較による検索を用いた検索システムでは安全に医療データを検索・閲覧することが可能である．

6.3　まとめ

シェア間の係数の差の比較による検索では，タグ t_1，t_2，しきい値 k と素数 p あるいは x_1，x_2 が漏洩した場合，kw が有限個に絞られる．これに対し，素数 p を大きくすれば問題ないが識別情報を有限個に絞り込ま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，p，X をシステムで秘匿することで安全性を担保する必要がある．

本章では，t_u，p，X を秘匿する検索システムの構成について述べた．提案システムでは t_u，p，X を秘匿することで提案検索方法の安全性を高めている．もし攻撃者がストレージにアクセスすることが可能であっても，提案検索方法を用いている場合シェアとタグの保存方法により，医療データを不正に復元される恐れはない．また，医師が使用する情報閲覧端末に攻撃者がアクセス可能であってもリモートアクセス先の情報端末の情報が入手できるのみで他の情報を入手することはできない．これより，シェア間の係数の差の比較による検索を用いた検索システムでは安全に医療データを検索・閲覧することが可能である．

第 7 章

おわりに

7.1 本研究のまとめ

本論文では部分復元可能な秘密分散法における検索の高速化を目的として，シェアの状態で検索を可能とする方法を提案した．さらに提案方法を用いるためのシェアとタグの保存方法を検討し，提案方法について速度と安全性の評価を行い，検索システムを提案した．

シェアの状態で検索を可能とする方法は，識別情報を $(2, n)$ しきい値秘密分散して作成したタグを医療データのシェアに付与する．そして検索キーワードをタグと同様に $(2, n)$ しきい値秘密分散して検索タグを作成し，検索タグと医療データのシェアに付与されているタグの係数の差を比較することでシェアの状態で検索を可能としている．タグと検索タグの係数の差を比較す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り，各ストレージで対応を持たせて保存しておく必要がある．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かってしまうため，医療データを不正復元される恐れがある．そこで同一の識別情報から作成した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は分からないようなシェアとタグの保存方法について検討した．これにより攻撃者がストレージにアクセスすることが可能であっても，シェアとタグの保存方法により医療データのシェアの組み合わせは分からず，タグからも医療データのシェアの組み合わせが分かることはないため，医療データを不正に復元される恐れはなくなる．

検索にかかる計算量は $\mathcal{O}(l^3)$(医療データのデータ長を l) から $\mathcal{O}(d^2)$(識別情報のデータ長を d) に削減でき，検索にかかる時間を計測したところ，目標である 70 万人分のデータを

15 秒以内に検索できることを確認した．また，提案方法では t_1，t_2，k と p あるいは x_1，x_2 が漏洩した場合には識別情報を絞り込まれるが，絞り込まれる数は p に依存しているため p を大きくすることで問題はない．しかし，識別情報が絞られることによって万が一医療データ不正に復元されてはならないことから，提案したシステム構成により t_u，p，X を秘匿することで安全性をより高められる．これより，シェア間の係数の差の比較による検索を用いた検索システムでは安全に医療データを入手することが可能である．

7.2　今後の課題

提案方法の検索にかかる時間は目標を達成することができたが，今回使用したデータはダミーデータであるため実際にデータが何個になるか定かではない．したがって，さらなる高速化が必要になる可能性がある．高速化の手法としては並列化や，有限体上で演算を行っているため拡大体に拡張することでさらなる高速化が見込めると考える．また，提案方法はしきい値 k が 2 で固定であるため $k > 2$ に拡張する必要がある．そして提案システムで秘匿したいデータが漏洩しないこと，ネットワーク越しでの検索時間などを明らかにするために，検索システムを作成し実験する必要がある．

謝辞

本研究を行うにあたり，ご指導いただきました高知工科大学情報学群の福本昌弘教授に謹んで感謝致します．本研究の副査をしていただいた情報学群敷田幹文教授，高田喜郎准教授，東京大学大学院情報理工学系研究科鵜川始陽准教授のお三方にも謹んで感謝致します．また，本研究で用いたデータの提供をいただいた高知医療センター情報システム室北村和之氏にも謹んで感謝致します．

福本先生には 4 年間という長い間大変お世話になりました．私の拙い日本語，乏しい理解力や適当にやり過ごそうとすることに対して呆れながらも何度もご指導いただいたき，何度か心がポキポキポキっといってしまいそうでしたが大変感謝しています．先生のおかげで求められていることは何か，どうすれば人に伝わるのかを学べたと思います．のびのびと研究をさせていただき，楽しく貴重な体験も経験させていただき，大変有意義な大学生活を送ることができました．

NOC の職員であり福本研究室の OB でもある福冨英次氏にも謹んで感謝致します．福冨氏には研究のアドバイス，発表スライドの添削やお食事に連れて行ってもらったりと公私共に大変お世話になりました．なかなか他では聞くことのできない話がたくさん聞けて楽しかったです．また，3 回ほど行われたつよつよ PC 組み立てでは，私の PC 組みたい欲を密かに満たさせてもらいました．株式会社ふくえい設立時や 13 年後？にあるであろう集まりには必ず馳せ参じます．

原田崇司助教にも謹んで感謝致します．原田先生には研究について一緒に悩んでいただいたり，お食事に連れて行ってもらったりと公私共に大変お世話になりました．神奈川に寄った際にはぜひご一報ください．また，居室と研究室にネットワークカメラを導入したという便りをお待ちしております．

研究室で共に研究を行ったり議論をしたはずの吉冨君，松本君，斎藤君にも謹んで感謝致します．同期のいない (某先生によると友達も少ないらしい) 寂しい私には，一緒にご飯を

謝辞

食べたり話をしてくれる貴重な存在でした．よく散歩に誘って研究の邪魔をしてすみませんでした．今後はちょこちょこ食事に誘って仕事の邪魔をしたいと思います．

福本研究室修士 1 年，3 年生の皆さん，たまにお話しして頂きありがとうございました．皆さんの謎のメンタルの強さがあれば何事でも乗り越えられると思います．

最後になりましたが，6 年間の大学生活を支えてくださった両親や兄弟，関わってくれた全ての皆様に感謝致します．

参考文献

[1] 福本昌弘，“高知県における電子カルテ遠隔バックアップと部分復元可能な秘密分散法，” 北隆館，Precision Medicine，pp. 57–61，Vol. 3，No. 9，2020.

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

[3] 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.

[4] 国立研究開発法人情報通信研究機構 (NICT)，“南海トラフ大地震に備えた高知の「地域医療情報バックアップ」を NICT の高セキュアな量子暗号でアシスト！–最先端の量子暗号化と秘密分散技術で安全に医療情報のセキュリティをアップ–”JGN インタビュー Vol. 007，https://testbed.nict.go.jp/jgn/ja/jgn-front/interview/007_1.html，2021 年 2 月 2 日閲覧.

[5] 高知県総務部統計分析課，“高知県の推計人口　月報 (令和 3 年 1 月 1 日現在)，”https://www.pref.kochi.lg.jp/soshiki/111901/files/2014021401751/r0301.pdf，2021 年 2 月 2 日閲覧.

[6] 小山博史，“電子カルテとは？，” 日本職業・災害医学会会誌，52，pp. 91–95，2004.

[7] 厚生労働省，“医療情報システムの安全管理に関するガイドライン第 5 版，”https://www.mhlw.go.jp/file/05-Shingikai-12601000-Seisakutoukatsukan-Sanjikanshitsu_Shakaihoshoutantou/0000166260.pdf，2017.

[8] 清水俊郎，“SS-MIX の現状 (SS-MIX2 の概要)，” http://www.jhit.jp/kanto_g01/6th20120908r.pdf，2012.

[9] 木村映善，松村泰志，三原直樹，黒田知宏，山下芳範，平松治彦，真鍋史郎，田中大介，佐藤敦，山倉直，“医療サービスの継続性を担保する電子カルテ秘密分散バックアップ技術の開発研究，”ICT イノベーションフォーラム 2015 予稿集，総務省，pp. 144–145，

Oct. 2015.

[10] 伊藤孝一，牛田芽生恵，山岡裕司，及川考徳，菊池浩明，“検索可能秘密分散方式の提案，”Vol. 2014–CSEC–64 No. 13.

[11] 中原将貴，岩村恵市，“サーバー台数を限定しない秘密分散法を用いた秘匿検索法の提案，”Vol. 2015–CSEC–69 No. 10.

[12] 吉冨亮平，“検索システムの構成による医療データ不正復元の防止，” 令和 2 年度高知工科大学学士学位論文.