平成 30 年度

学士学位論文

部分復元可能な秘密分散法における秘密分散データを用いた検索

Medical Record Retrieval in Confined Decodable Secret Sharing Scheme Using Secret Sharing Data

1190350　　中村　巴

指導教員　　福本　昌弘

2019 年 3 月 15 日

高知工科大学 情報学群

要 旨

部分復元可能な秘密分散法における秘密分散データを用いた検索

中村 巴

災害時，外部から派遣された医療従事者に医療データを提供することができれば，被災地での医療行為を円滑に行うことができる．また，災害時は，ネットワークや電源などのリソースが十分に用意できないため，医療行為に最低限必要なデータを提供できると良い．そこで，部分復元可能な秘密分散システムが提案された．医師は診療を行う際，患者の医療データを分散バックアップデータから検索して閲覧する必要がある．しかし，分散バックアップデータをシェアの状態で検索することはできないため，名前などの情報を部分的に復元してから検索を行う必要がある．しかし，部分復元は演算量が大きく，復元速度が遅いため，速く検索することはできない．そこで，部分復元可能な秘密分散バックアップした医療データ検索方法が提案された．しかし，検索のために付与した情報を復元するため，分散バックアップデータの数に比例して検索にかかる演算量が大きくなる．

本論文では，検索の高速化を目的として，検索のために付与した情報を復元せず，シェアの状態で検索可能な手法を提案し，従来の検索方法と提案検索方法の比較を行っている．提案検索方法では，検索用の情報と検索キーワードの分散に定数を用いることでシェア同士での一致の判定を可能とし，検索にかかる演算量を小さくしている．従来の検索方法と提案検索方法を比較した結果，検索にかかる時間を短縮できていることを明らかにしているが，定数を用いたことによって従来の検索方法より安全性が劣ることも明らかにしている．

キーワード　秘密分散法，検索

Abstract

Medical Record Retrieval in Confined Decodable Secret Sharing Scheme Using Secret Sharing Data

Tomo NAKAMURA

A Disaster Medical Assistance Team (DMAT) is a rescue team for natural disaster situations. In order to optimize their rescue performance, DMATs have to request information of curing patients, such as blood type or medication history from data center. Regarding network traffic on the disaster situation, it is difficult to provide all information to DMATs for a short time. To overcome this difficulty, confined decodable secret sharing scheme system was proposed since 2015. In the previous system, patient's medical data can be retrieved by decoding information of patient's name; all this can result in large amount of computation and not suitable on the disaster situation. Retrieval method for medical data by confined decodable secret sharing scheme have been proposed for reduce this computational complexity, however, this method still requires computation time since it has to decode all retrieval information.

This paper therefore proposes retrieval method that does not require decode of retrieval information. In the proposed method, the retrieval information and retrieval keyword are distributed using Shamir's secret sharing scheme with fixed high-ordered polynomial coefficients or initial random constants, then the data can be retrieved by matching between the shares and the keyword without decoding any shared information. As a result of comparing the previous retrieval method and proposed method, it is clarified that the time required for the retrieval can be shortened, and it is also

demonstrated that the safety is inferior than the previous method caused by the initial random constants, which are no safety concerns in this study.

key words secret sharing scheme, retrieval

目次

図目次

表目次

第 1 章

序論

1.1 本研究の背景と目的

災害時，外部から派遣された医療従事者に医療データを提供することができれば，被災地での医療行為を円滑に行うことができる．また，災害時は，ネットワークや電源などのリソースが十分用意できないため，医療行為に最低限必要なデータを提供できると良い．そこで，部分復元可能な秘密分散システムが提案された [1]．医師は診療を行う際，患者の医療データを分散バックアップデータから検索して閲覧する必要がある．しかし，提案されたシステムは，分散バックアップデータをシェアの状態で検索することは出来ない．そのため，医療データを検索したい場合，名前などの情報を部分的に復元してから医療データを検索する必要がある．部分復元可能な秘密分散システムは復元にかかる演算量が大きく，検索に時間がかかってしまう．そこで，速く検索を行うために，部分復元可能な秘密分散バックアップした医療データ検索方法が提案された [2]．提案された検索方法は，分散バックアップデータに検索用の情報を秘密分散して付与する．医療データを検索したい場合には，付与した検索用の情報だけを復元し，検索を行う．また，検索用の情報のデータサイズを削減し，検索にかかる演算量を小さくした．しかし，データサイズが小さいとはいえ，検索用の情報を復元するため，分散バックアップデータの数に比例して演算量は大きくなり，検索に時間がかかる．

本研究では，検索の高速化を目的とし，シェア同士で一致の判定を可能とする方法を提案する．そして，部分復元可能な秘密分散バックアップした医療データ検索方法と提案検索方法の比較を行う．

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，(k, n) しきい値秘密分散法と部分復元可能な秘密分散法について述べた後，部分復元可能な秘密分散バックアップした医療データ検索方法と検索にかかる演算量について述べ，付与する検索用の情報とデータサイズ削減について述べる．3 章では，秘密分散データ同士で一致の判定を可能とする方法について述べた後，秘密分散データを用いて検索可能であることの証明を述べ，部分復元可能な秘密分散法における秘密分散データを用いた検索を提案し，検索にかかる演算量について述べ，従来の検索方法との比較を述べる．4 章では従来の検索方法と秘密分散データを用いた検索の実装について述べる．5 章では，従来の検索方法と秘密分散データを用いた検索の検索にかかる時間を計測し，比較する．6 章では，本研究をまとめ，今後の課題を述べる．

第 2 章

部分復元可能な秘密分散システムと検索方法

東日本大震災で津波によってカルテなどの医療データが紛失したことにより，被災地での医療行為に支障が出た．そこで，高知県では大規模災害に備え，医療データを遠隔地にバックアップする取り組みが行われている．また，バックアップした医療データを被災地での医療行為に活用することが期待されている [3]．災害時は，負傷者の増加と医師の不足が想定されるため，被災地域外から派遣された外部の医師も患者の治療を行う．そのため，外部の医師にも治療に必要な医療データを提供することができれば，適切で円滑な治療を行うことができる．そこで，部分復元可能な秘密分散システムが提案された [1]．しかし，提案されたシステムは，分散バックアップデータをシェアの状態で検索することは出来ない．そのため，医療データを検索したい場合，名前などの情報を部分的に復元してから医療データを検索する必要がある．部分復元可能な秘密分散システムは復元にかかる演算量が大きく，検索に時間がかかってしまう．そこで，速く検索を行うために，部分復元可能な秘密分散バックアップした医療データ検索方法が提案された [2]．提案された方法は，分散バックアップデータに検索用の情報を秘密分散して付与する．医療データを検索したい場合には，付与した情報だけを復元し，検索を行う．本章では，部分復元可能な秘密分散システムに用いられている分散バックアップの手法である (k, n) しきい値秘密分散法について述べ，部分復元可能な秘密分散システムについて述べる．そして部分復元可能な秘密分散バックアップした医療データ検索方法と検索にかかる演算量について述べる．

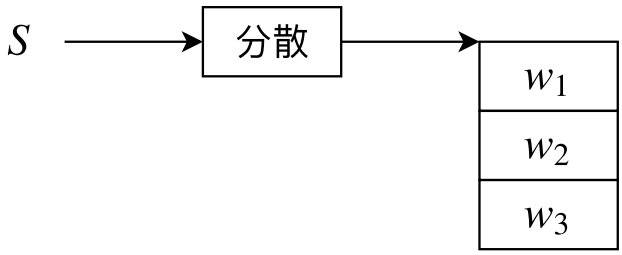

図 2.1 (k, n) しきい値秘密分散法の分散の流れ

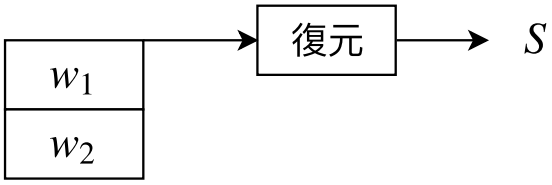

図 2.2 (k, n) しきい値秘密分散法の復元の流れ

2.1 (k, n) しきい値秘密分散法 [4]

部分復元可能な秘密分散システムで用いられている分散バックアップの手法である (k, n) しきい値秘密分散法について述べる．(k, n) しきい値秘密分散法の分散の流れを図 2.1，復元の流れを図 2.2 に示す． 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復元可能な方法である．以下に，(k, n) しきい値秘密分散法の手法を示す.

分散したいデータを S とする．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

$$X = \{x_1, x_2, \cdots, x_n\}$$

とおくことができる．X より

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

2.1　(k, n) しきい値秘密分散法 [4]

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

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

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

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

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，分散する．

復元時は，シェアを k 個集める．集めたシェアからシェア行列 $\mathbf{W}$

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

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

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

となり，データ S を復元できる．

2.2　部分復元可能な秘密分散システム

秘密分散法は分散したデータを全て復元できるかでき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．災害時はネットワークや電源などのリソースが十分に用意できないため，治療に最低限必要な医療データを閲覧できればよい．そこで，部分復元可能な秘密分散システムが提案された [1]．本節では，部分復元アルゴリズムについて述べる．部分復元アルゴリズムは，分散段階と復元段階の 2 段階からなる．

2.2.1　分散段階

分散段階の流れを図 2.3 に示す．まず，分散しようとするデータ S を意味のある項目ごとに e 個に分け，分けたデータ $S_i(i = 1, 2, \cdots, e)$ それぞれに対して (k, n) しきい値秘密分散法を用いてシェアを生成し分散する．データを意味のある項目ごとに分けることを分割とす

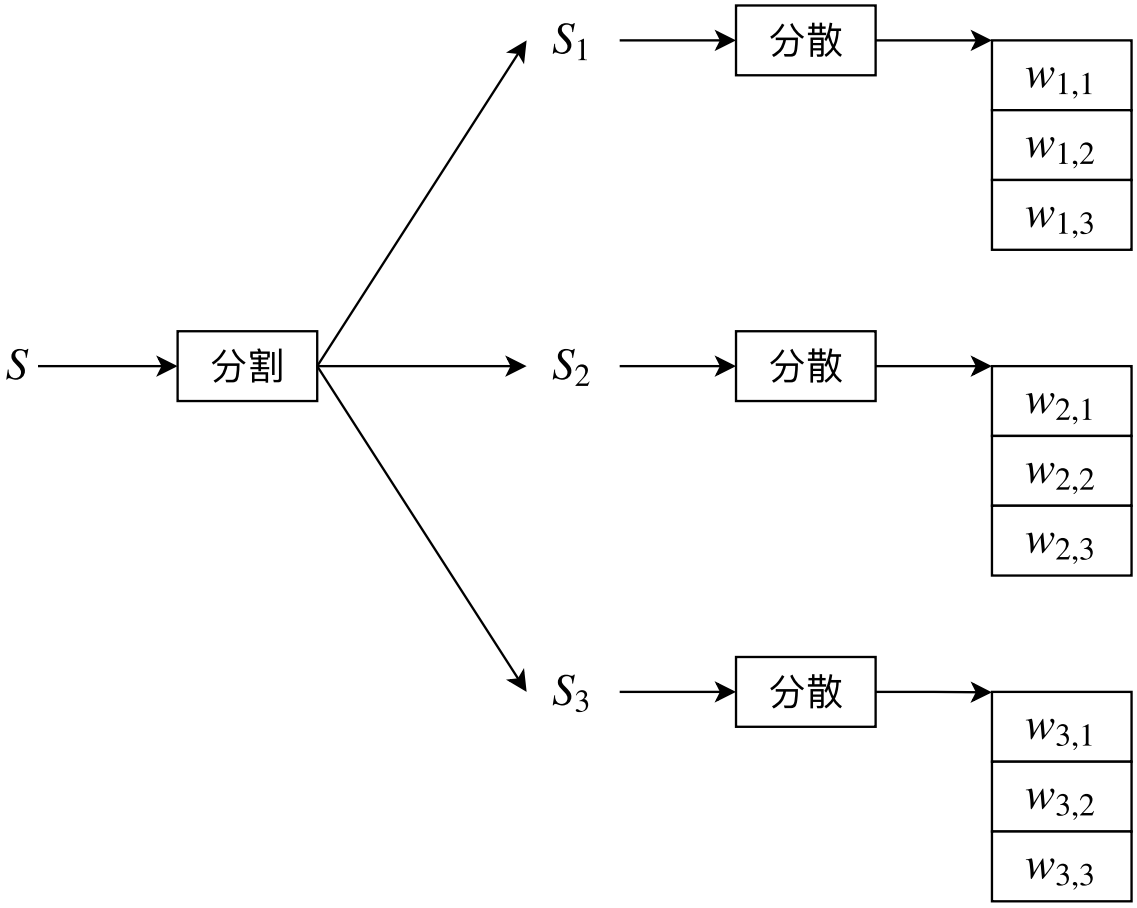

図 2.3　部分復元可能な秘密分散システムの分散段階の流れ

る．シェアの状態ではどのシェアがどの項目のシェアであるかの判別は困難であるため，項目が正しい順番で並んだデータを復元できるように，正しく紐付ける情報を作成し紐付け情報とする．紐付け情報は分割したデータのサイズ l_i を要素として対応する項目ごとに並べたものであり，復元したくない項目に対応する要素を 0 となるように作成する．

2.2.2　復元段階

復元段階の流れを図 2.4 に示す．分散したシェアから復元したいシェアのみをしきい値以上集め，復元したい項目のみを含んだデータのシェアとなるように紐付け情報を用いて結合

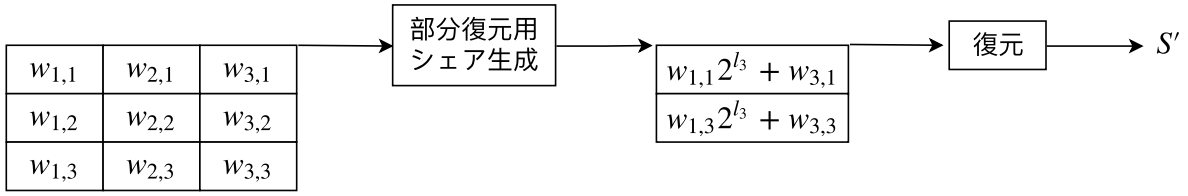

図 2.4　部分復元可能な秘密分散システムの復元段階の流れ

し，部分復元用シェアを作成する．部分復元用シェアを (k, n) しきい値秘密分散法の復元操作を行うことで，復元したい項目のみを含んだデータを復元することができる．

2.2.3　部分復元の手順

分散しようとするデータを S とする．S を意味のある項目ごとに分割したデータを $S_i(i = 1, 2, \cdots, e)$ とする．分割したデータのデータサイズを l_i とする．データ S は分割したデータ S_i を用いて

$$S = S_1 \prod_{m=2}^{e} 2^{l_m} + S_2 \prod_{m=3}^{e} 2^{l_m} + \cdots + S_{e-1}2^{l_e} + S_e \tag{2.1}$$

と表すことができる．分割データ S_i の集合を

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

とおき，復元したいデータの集合を S_{all} の部分集合として

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

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

$$\boldsymbol{u}_c = \begin{pmatrix} \prod_{m=2}^{j} 2^{l_m} \\ \prod_{m=3}^{j} 2^{l_m} \\ \vdots \\ 1 \end{pmatrix} \tag{2.2}$$

を作成する．部分復元の詳細な手順を以下に示す．

1. S を (2.1) 式を用いて分割し，$S_i (i = 1, 2, \cdots, e)$ を作成する．S_i のデータサイズを l_i とする．
2. 素数 $p (S < p$ かつ $n < p)$ を選択する．
3. $\mathbb{Z}/p\mathbb{Z} - \{0\}$ の集合 $X = \{x_1, x_2, \cdots, x_n\}$ 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 $\mathbf{A}_i$

$$\mathbf{A}_i = \begin{pmatrix} S_i \\ r_{i,1} \\ r_{i,2} \\ \vdots \\ r_{i,k-1} \end{pmatrix}$$

を作成する．

4. 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．$w_{i,j}$ を分散する．

5. G_c から各 w_{c_j} のシェアが k 個になるように集める．k 個のシェアからシェア行列 $\mathbf{W}_c$

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

を作成する．

6. $\mathbf{W}_c$ に対して (2.2) 式をかけ，部分復元用シェア $\mathbf{W}_c\boldsymbol{u}_c$ を作成する．

$$\begin{aligned} \mathbf{W}_c\boldsymbol{u}_c &= \begin{pmatrix} w_{c_{1,1}} & w_{c_{2,1}} & \cdots & w_{c_{j,1}} \\ w_{c_{1,2}} & w_{c_{2,2}} & \cdots & w_{c_{j,2}} \\ \vdots & \vdots & \ddots & \vdots \\ w_{c_{1,k}} & w_{c_{2,k}} & \cdots & w_{c_{j,k}} \end{pmatrix} \begin{pmatrix} \prod_{m=2}^{j} 2^{l_m} \\ \prod_{m=3}^{j} 2^{l_m} \\ \vdots \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} w_{c_{1,1}} \prod_{m=2}^{j} 2^{l_m} + w_{c_{2,\,1}} \prod_{m=3}^{j} 2^{l_m} + \cdots + w_{c_{j,1}} \\ w_{c_{1,2}} \prod_{m=2}^{j} 2^{l_m} + w_{c_{2,\,2}} \prod_{m=3}^{j} 2^{l_m} + \cdots + w_{c_{j,2}} \\ \vdots \\ w_{c_{1,k}} \prod_{m=2}^{j} 2^{l_m} + w_{c_{2,\,k}} \prod_{m=3}^{j} 2^{l_m} + \cdots + w_{c_{j,k}} \end{pmatrix} \end{aligned} \tag{2.3}$$

(2.3) 式を展開してまとめると,

$$\mathbf{W}_c\boldsymbol{u}_c = \mathbf{X} \begin{pmatrix} S_{c_1} \prod_{m=2}^{j} 2^{l_m} + S_{c_2} \prod_{m=3}^{j} 2^{l_m} + \cdots + S_{c_j} \\ r_{1,1} \prod_{m=2}^{j} 2^{l_m} + r_{2,1} \prod_{m=3}^{j} 2^{l_m} + \cdots + r_{j,1} \\ \vdots \\ r_{1,k-1} \prod_{m=2}^{j} 2^{l_m} + r_{2,k-1} \prod_{m=3}^{j} 2^{l_m} + \cdots + r_{j,k-1} \end{pmatrix} \tag{2.4}$$

となる．

7. (2.4) 式に $\mathbf{X}$ の逆行列 $\mathbf{X}^{-1}$ を左からかけると

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

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．

2.3　部分復元可能な秘密分散バックアップした医療データ検索方法

部分復元可能な秘密分散システムは，シェアの状態で検索できないため，医療データを検索したい場合，名前などの情報を部分的に復元してから医療データを検索する必要がある．しかし，部分復元可能な秘密分散システムは演算量が大きく，復元速度が遅い．そこで，速く検索を行うために，部分復元可能な秘密分散バックアップした医療データ検索方法が提案された [2]．本節では，部分復元可能な秘密分散バックアップした医療データ検索方法と検索にかかる演算量について述べる．そして，付与する検索用の情報について述べ，検索用の情報のサイズの削減と検索方法について述べる．部分復元可能な秘密分散バックアップした医療データ検索方法は，検索情報を付与する分散段階と，検索段階の 2 段階からなる．

2.3.1　分散段階

医療データを S とする．S の分散の流れを図 2.5，検索用の情報の分散の流れを図 2.6，検索用の情報の付与の流れを図 2.7 に示す． S を部分復元可能な秘密分散してシェアを生成する．S の一部の情報を利用し，検索用の情報 RI を生成する．素数を $p(RI < p$ 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 とおく．すなわち X は

$$X = \{x_1, x_2, \cdots, x_n\}$$

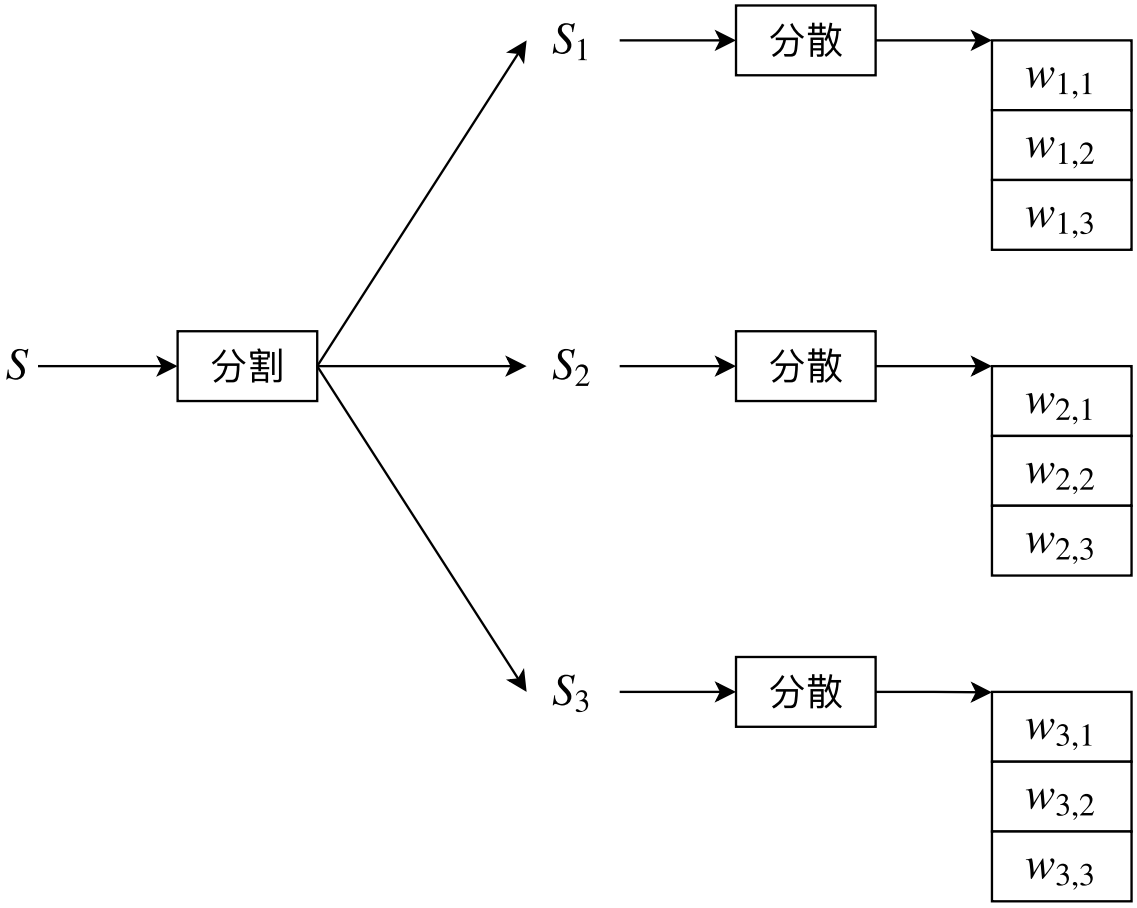

図 2.5　医療データ検索方法の医療データの分散の流れ

とおくことができる．X から

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

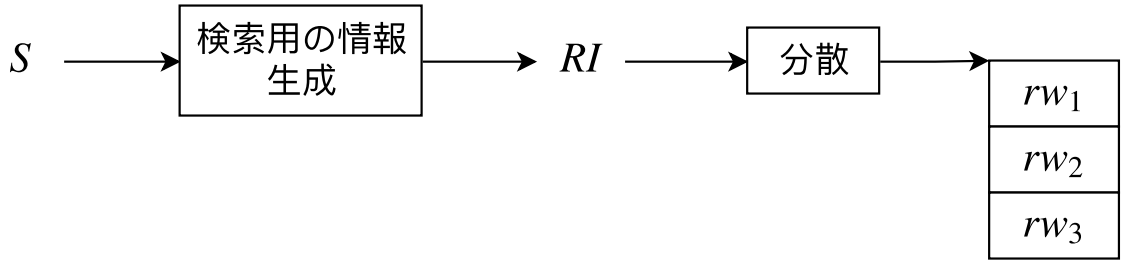

図 2.6　医療データ検索方法の検索用情報の分散の流れ

となるような vandermonde 行列 $\mathbf{X}$ を作成する．RI と R から $k \times 1$ の行列 $\mathbf{RA}$

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

を作成する．p を法とした行列 $\mathbf{X}$ と $\mathbf{RA}$ の乗算から

$$\mathbf{XRA} = \begin{pmatrix} rw_1 \\ rw_2 \\ \vdots \\ rw_n \end{pmatrix} \pmod{p}$$

となるような検索用シェア $rw_i (i = 1, 2, \cdots, n)$ が得られる．検索用シェアを S のシェアに付与する．

2.3.2　検索段階

検索段階の流れを図 2.8 に示す．検索キーワードを RK とする．ある医療データのシェア

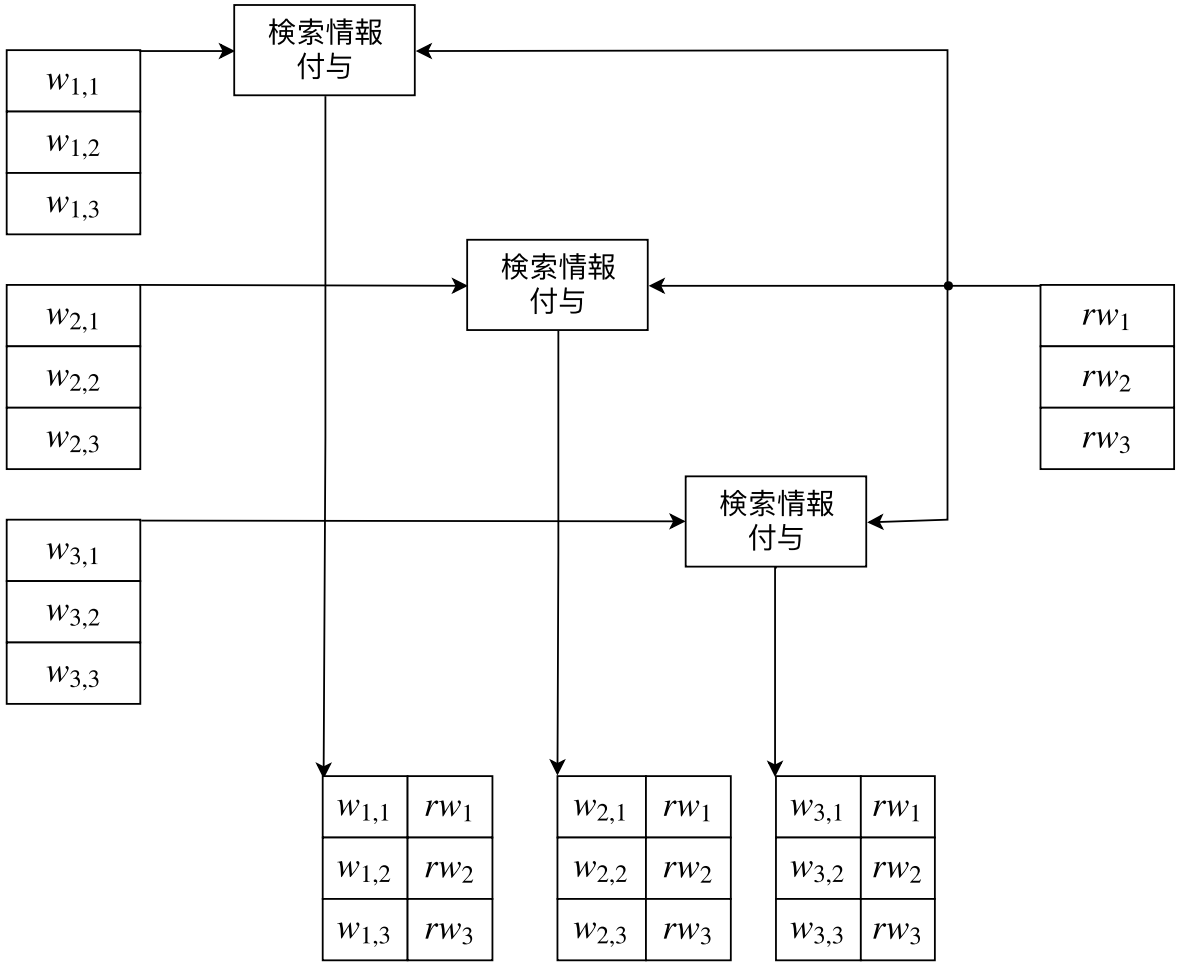

図 2.7　医療データ検索方法の検索用情報の付与の流れ

に付与されている検索用シェアを k 個集めて検索用シェア行列 **RW**

$$\mathbf{RW} = \begin{pmatrix} rw_1 \\ rw_2 \\ \vdots \\ rw_k \end{pmatrix}$$

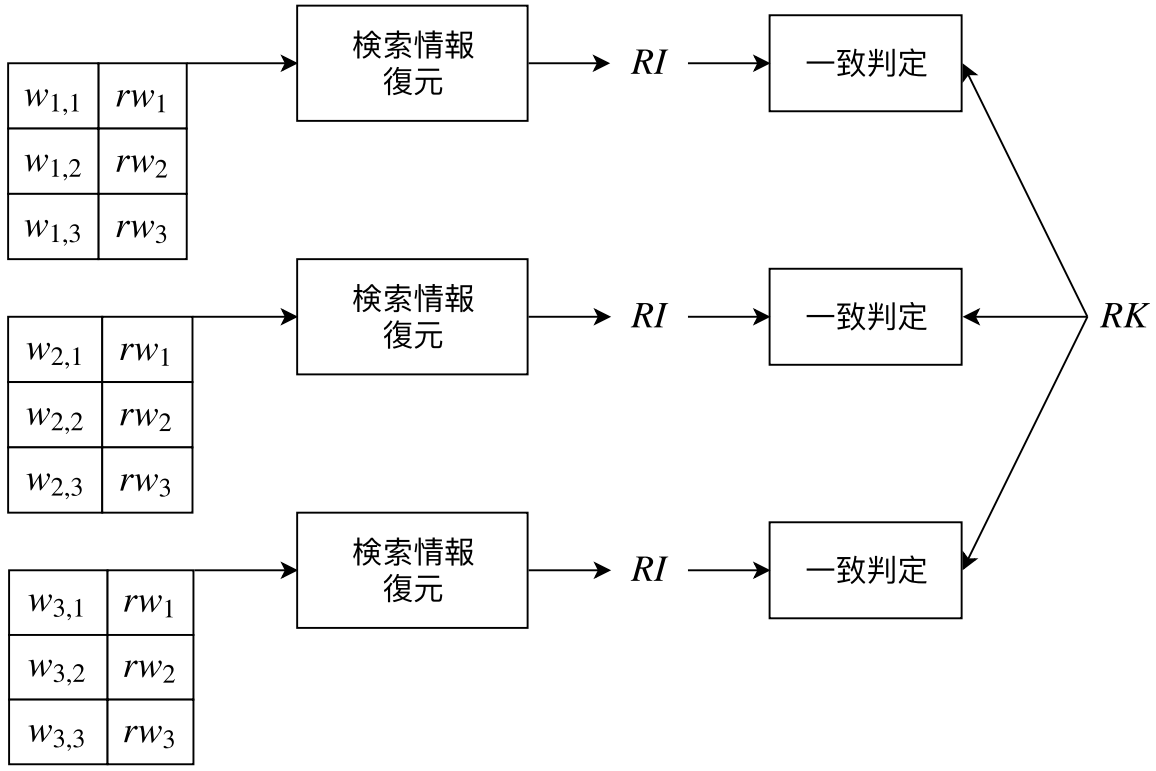

図 2.8　医療データ検索方法の検索の流れ

を作成する．

RW に **X** の逆行列 $\mathbf{X}^{-1}$ を左からかけると

$$\mathbf{X}^{-1}\mathbf{RW} = \begin{pmatrix} RI \\ r_1 \\ r_2 \\ \vdots \\ r_{k-1} \end{pmatrix}$$

となり，検索用の情報 RI を復元するすることができる．RI と RK の一致判定を行う．以上の操作をすべての医療データに付与されている検索用シェアに対して行い，該当するデータを見つける．

2.3.3　検索にかかる演算量

検索にかかる演算量を検索手順に沿って示す．計算コストの高い乗算演算に着目して演算量を求める．

検索キーワード RK を検索したいとする．ある医療データのシェアに付与されている検索用シェアを k 個集め，検索用シェア行列 $\mathbf{RW}$

$$\mathbf{RW} = \begin{pmatrix} rw_1 \\ rw_2 \\ \vdots \\ rw_k \end{pmatrix}$$

を作成する．集めた検索用シェアに対応した $x_i(i = 1, 2, \cdots, k)$ を用いて $k \times k$ の vandermonde 行列 $\mathbf{X}$

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

を作成する．vandermonde 行列 $\mathbf{X}$ の作成に必要な乗算回数 V は,

$$V = \begin{cases} 0 & (k < 3) \\ k \sum_{c=1}^{k-2} c & (otherwise) \end{cases}$$

回となる．$k \times k$ の $\mathbf{X}$ の逆行列 $\mathbf{X}^{-1}$ を掃き出し法を用いて作成する．逆行列 $\mathbf{X}^{-1}$ の作成に必要な乗算回数 R は,

$$R = k^3$$

回となる．$\mathbf{RW}$ に対して $\mathbf{X}^{-1}$ を左からかけると,

$$\mathbf{X}^{-1}\mathbf{RW} = \begin{pmatrix} RI \\ r_1 \\ r_2 \\ \vdots \\ r_{k-1} \end{pmatrix}$$

となり，検索用の情報 RI を復元することができる．$\mathbf{X}^{-1}\mathbf{RW}$ に必要な乗算回数 M は

$$M = k^2$$

回となる．以上の手順より，検索用の情報 1 個の復元にかかる合計乗算回数は $V + R + M$ となり，

$$\begin{cases} k^3 + k^2 & (k < 3) \\ k^3 + k^2 + k\sum_{c=1}^{k-2} c & (otherwise) \end{cases} \tag{2.5}$$

回となる．そして，RI と RK の一致判定を行う．以上の操作を医療データに付与されているすべての検索用の情報に対して行うため，医療データのシェアの総数を Y とすると，検索にかかる合計乗算回数は (2.5) 式より

$$\begin{cases} Y(k^3 + k^2) & (k < 3) \\ Y(k^3 + k^2 + k\sum_{c=1}^{k-2} c) & (otherwise) \end{cases} \tag{2.6}$$

回となる．検索用の情報 RI のデータサイズを d とし，1 回の乗算にかかる乗算の演算量を d^2 とすると，検索にかかる演算量は，(2.6) 式より

$$\begin{cases} d^2Y(k^3 + k^2) & (k < 3) \\ d^2Y(k^3 + k^2 + k\sum_{c=1}^{k-2} c) & (otherwise) \end{cases}$$

となる．

2.3.4　付与する検索用の情報

部分復元可能な秘密分散バックアップした医療データ検索方法では，検索用の情報を付与し，検索時は検索用の情報を復元してから検索する．そのため，検索用として付与する患者情報を定める．付与する患者情報の条件を以下に示す．

- 医療データから得られる情報である
- 患者全員が持っている情報である

表 2.1　付与する検索用の情報

患者姓	患者名	性別	西暦	月日	住所	電話番号

表 2.2　姓を検索した場合の付与文字数ごとの中央値と最大人数，最少人数

付与文字数	1 文字	2 文字	3 文字	4 文字
中央値 (人)	29517	12939	8068	8068
最大人数 (人)	117776	61969	54689	54689
最少人数 (人)	5701	5701	5701	5701

- 災害時に入手できる可能性のある情報である

以上の条件を満たす情報を表 2.1 に示す．患者の外見から判断できるのは性別，西暦であるが，この 2 つを組み合わせて検索しても患者データを特定することはできない．また，他の患者情報について考えると，患者姓を聞き出せない場合は，他の情報を聞き出せない可能性が高い．よって，患者姓を検索用の情報として付与した場合を考える．

2.3.5　検索用の情報のサイズ削減

部分復元可能な秘密分散バックアップした医療データ検索方法は，付与した検索用の情報を復元してから検索を行う．災害急性期は，患者データを速く検索する必要がある．しかし，検索用の情報のデータサイズが大きければ，復元処理の演算量が大きくなり時間がかかってしまう．そこで，検索用の情報のデータサイズの削減を行い，復元処理にかかる演算量を小さくする．まず，姓の読み先頭何文字でどのくらいの人数がヒットするのかを示す．日本人の性と名の分布 [5] から東邦生命の被保険者と保険者の上位 60 位の姓の読みを利用し，744849 人のうち，付与文字数ごとの中央値とヒットする最大，最小の人数を表 2.2 に示す．表 2.2 より，3 文字付与すれば 1 つの姓が特定できることから，姓の読み先頭 3 文字を検索用の情報とすれば良いことが分かる．

表 2.3　検索手順による最大復元回数と演算量

	1 文字ずつ	1 文字，2 文字	2 文字，1 文字	3 文字
最大復元回数 (回)	924594	862625	806818	744849
演算量	9,231,146,496	12,140,074,752	30,364,988,160	66,929,151,744

2.3.6　検索方法

付与した姓を分散し検索を行うには 4 通りの検索方法がある．1 つ目は，1 文字ずつ分割し，分散して 1 文字ずつ復元し検索を行う方法．2 つ目は 1 文字，2 文字に分割し，分散して 1 文字復元し検索を行った後，2 文字復元し検索を行う方法．3 つ目は 2 文字，1 文字に分割し，分散して 2 文字復元し検索を行った後，1 文字復元し検索を行う方法．4 つ目は，3 文字まとめて分散して 3 文字復元し検索を行う方法である．以上の 4 通りの方法の最大復元回数を表 2.2 より求め，演算量を (2.6) 式を用いて，表 2.3 に示す．なお，秘密分散法は (3,5) しきい値とし，平仮名 1 文字のデータサイズを 2bytes とする．表 2.3 より，姓を 1 文字ずつ分散して，1 文字ずつ復元し検索を行うのが良いことが分かる．

2.3.7　部分復元可能な秘密分散バックアップした医療データ検索方法の評価

部分復元可能な秘密分散バックアップした医療データ検索方法は，検索用の情報を復元してから検索を行う．しかし，分散バックアップデータの数に比例して演算量は大きくなり，検索に時間がかかる．そこで，シェア同士で一致判定を行うことが可能であれば，検索用の情報を復元する必要がないため，検索にかかる時間の短縮が期待できる．しかし，部分復元可能な秘密分散バックアップした医療データ検索方法は，検索用の情報の分散に乱数を用いており，乱数は復元に必要ないため記憶しない．すると，検索用の情報と同じ検索キーワードが入力されても，検索用シェアと同じシェアを生成するのは困難である．また，乱数を用

いているため，異なる検索用の情報から同じ検索用シェアが生成される可能性がある．以上より，部分復元可能な秘密分散バックアップした医療データ検索方法は，シェア同士で一致判定を行うことができず，検索用の情報を復元する必要がある．

本研究では，シェア同士での一致の判定を可能とする方法を提案し，復元処理が不要な検索を実現する．

第 3 章

部分復元可能な秘密分散法における秘密分散データを用いた検索

部分復元可能な秘密分散バックアップした医療データ検索方法が提案された．提案された方法は，秘密分散した医療データに，小さいデータサイズの検索用の情報を秘密分散して作成した検索用シェアを付与し，検索時には検索用シェアを復元することで検索を行う．しかし，データサイズが小さいとはいえ，検索用の情報の復元を行うため，検索には大量の演算量を必要とする．そこで，速く検索を行うために，秘密分散データを用いた復元不要な検索を提案する．本章では，秘密分散データ同士での一致の判定を可能とする方法について述べ，提案方法を用いて検索可能であることの証明を行う．次に，秘密分散データを用いた検索について述べ，検索にかかる演算量を述べる．

3.1　秘密分散データ同士での一致の判定を可能とする方法

部分復元可能な秘密分散システムにおける検索を高速化するために，秘密分散データ同士での一致の判定を可能とする方法を提案する．

3.1.1　基本的な考え

部分復元可能な秘密分散バックアップした医療データ検索方法では，検索用の情報を復元してから検索を行う必要がある．そこで，シェア同士で一致判定を行うことができれば，検索における復元処理を省略することができ，検索を高速化することができるという考えから

秘密分散データ同士での一致の判定を可能とする方法を提案する．

3.1.2　提案方法

検索用の情報 RI，検索キーワード RK を考える．素数を $p(RI,\, RK < p$ かつ $n < p)$，$\mathbb{Z}/p\mathbb{Z} - \{0\}$ 上の集合を X，$\mathbb{Z}/p\mathbb{Z}$ 上の集合を R とおく．すなわち X は

$$X = \{x_1,\, x_2, \cdots,\, x_n\}$$

とおくことができる．X 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 vandermonde 行列 $\mathbf{X}$ を作成する．RI, RK と R から $k \times 1$ の行列 $\mathbf{RA}$, $\mathbf{KA}$

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

を作成する．p を法とした行列 $\mathbf{X}$ と $\mathbf{RA}$, $\mathbf{KA}$ の乗算から

$$\mathbf{XRA} = \begin{pmatrix} rw_1 \\ rw_2 \\ \vdots \\ rw_n \end{pmatrix} \pmod{p},\ \mathbf{XKA} = \begin{pmatrix} kw_1 \\ kw_2 \\ \vdots \\ kw_n \end{pmatrix} \pmod{p}$$

となるようなシェア rw_i, $kw_i (i = 1,\, 2, \cdots,\, n)$ が得られる．これらを方程式で表すと

$$rw_i = RI + r_1 x_i + r_2 x_i^2 + \cdots + r_{k-1} x_i^{k-1} \pmod{p} \tag{3.1}$$

$$kw_i = RK + r_1 x_i + r_2 x_i^2 + \cdots + r_{k-1} x_i^{k-1} \pmod{p} \tag{3.2}$$

となる．(3.1) 式と (3.2) 式より，$RI = RK$ であれば，$rw_i = kw_i$ であることがわかる．

このように，分散に共通の集合 X, R を用いることで，秘密分散データ同士での一致の判定を可能とする．

3.2　検索が可能であることの証明

本節では，秘密分散データ同士での一致の判定を可能とする方法を用いて検索が可能であることの証明を行う．秘密分散データを用いた検索が可能であるためには，検索用の情報と検索キーワードが同じである場合に生成される検索用シェアと検索キーシェアが一致し，異なる検索用の情報から生成した検索用シェアは同じ値になってはいけない．また，検索キーシェアも同様に，異なる検索キーワードから生成した検索キーシェアが同じ値になってはいけない．したがって，以下の条件を満たす必要がある．

- 検索用シェアと検索キーシェアが一致する
- 検索用シェアと検索キーシェアがそれぞれ一意である

3.2.1　検索用シェアと検索キーシェアが一致する

検索用の情報を RI とする．素数を $p(RI < p$ かつ $n < p)$，$\mathbb{Z}/p\mathbb{Z} - \{0\}$ 上の集合を X，$\mathbb{Z}/p\mathbb{Z}$ 上の集合を R とおく．すなわち X は

$$X = \{x_1, x_2, \cdots, x_n\}$$

とおくことができる．X 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 vandermonde 行列 $\mathbf{X}$ を作成する．RI と R から $k \times 1$ の行列 $\mathbf{RA}$

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

を作成する．p を法とした行列 $\mathbf{X}$ と $\mathbf{RA}$ の乗算から

$$\mathbf{XRA} = \begin{pmatrix} rw_1 \\ rw_2 \\ \vdots \\ rw_n \end{pmatrix} \pmod{p}$$

となるようなシェア $rw_i (i = 1, 2, \cdots, n)$ が得られる．これを連立方程式で表すと

$$\begin{cases} rw_1 = RI + r_1 x_1 + r_2 x_1^2 + \cdots + r_{k-1} x_1^{k-1} \pmod{p} \\ rw_2 = RI + r_1 x_2 + r_2 x_2^2 + \cdots + r_{k-1} x_2^{k-1} \pmod{p} \\ \vdots \\ rw_n = RI + r_1 x_n + r_2 x_n^2 + \cdots + r_{k-1} x_n^{k-1} \pmod{p} \end{cases} \tag{3.3}$$

となる．

検索キーワードを RK とする．分散段階で用いた集合 X，R と RK 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 $\mathbf{X}$ と，$k \times 1$ の行列 $\mathbf{KA}$

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

を作成する．p を法とした行列 $\mathbf{X}$ と $\mathbf{KA}$ の乗算から

$$\mathbf{XKA} = \begin{pmatrix} kw_1 \\ kw_2 \\ \vdots \\ kw_n \end{pmatrix} \pmod{p}$$

となるような検索キーシェア kw_i が得られる．これを連立方程式で表すと

$$\begin{cases} kw_1 = RK + r_1x_1 + r_2x_1^2 + \cdots + r_{k-1}x_1^{k-1} \pmod p \\ kw_2 = RK + r_1x_2 + r_2x_2^2 + \cdots + r_{k-1}x_2^{k-1} \pmod p \\ \qquad\qquad \vdots \\ kw_n = RK + r_1x_n + r_2x_n^2 + \cdots + r_{k-1}x_n^{k-1} \pmod p \end{cases} \tag{3.4}$$

となる．(3.3) 式と (3.4) 式より，$RI = RK$ であれば一致することがわかる．したがって，検索用の情報と検索キーワードが同じである場合，作成される検索用シェアと検索キーシェアは一致する．

3.2.2　検索用シェアと検索キーシェアがそれぞれ一意である

検索用の情報の集合を $\mathbf{RI} = \{RI_1, RI_2, \cdots, RI_m\}$ とする．ある検索用の情報 RI_g , $RI_h(\in \mathbf{RI})$ を考える．素数を $p(RI_g, RI_h < p$ かつ $n < p)$ とし，$\mathbb{Z}/p\mathbb{Z} - \{0\}$ 上の集合を X，$\mathbb{Z}/p\mathbb{Z}$ 上の集合を R とおく．すなわち X は

$$X = \{x_1, x_2, \cdots, x_n\}$$

とおくことができる．X 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 vandermonde 行列 $\mathbf{X}$ を作成する．RI_g, RI_h と R から $k \times 1$ の行列 $\mathbf{RA}_g$, $\mathbf{RA}_h$

$$\mathbf{RA}_g = \begin{pmatrix} RI_g \\ r_1 \\ r_2 \\ \vdots \\ r_{k-1} \end{pmatrix}, \mathbf{RA}_h = \begin{pmatrix} RI_h \\ r_1 \\ r_2 \\ \vdots \\ r_{k-1} \end{pmatrix} \quad (r_1, r_2, \cdots, r_{k-1} \in R, r_{k-1} \neq 0)$$

を作成する．p を法とした行列 $\mathbf{X}$ と $\mathbf{RA}_g$, $\mathbf{RA}_h$ の乗算から

$$\mathbf{XRA}_g = \begin{pmatrix} rw_{g,1} \\ rw_{g,2} \\ \vdots \\ rw_{g,n} \end{pmatrix} \pmod p,\ \mathbf{XRA}_h = \begin{pmatrix} rw_{h,1} \\ rw_{h,2} \\ \vdots \\ rw_{h,n} \end{pmatrix} \pmod p$$

となるような検索キーシェア $rw_{g,i}$, $rw_{h,i} (i = 1, 2, \cdots, n)$ が得られる．これらを連立方程式で表すと

$$\begin{cases} rw_{g,1} = RI_g + r_1 x_1 + r_2 x_1^2 + \cdots + r_{k-1} x_1^{k-1} \pmod p \\ rw_{g,2} = RI_g + r_1 x_2 + r_2 x_2^2 + \cdots + r_{k-1} x_2^{k-1} \pmod p \\ \vdots \\ rw_{g,n} = RI_g + r_1 x_n + r_2 x_n^2 + \cdots + r_{k-1} x_n^{k-1} \pmod p \end{cases}$$

$$\begin{cases} rw_{h,1} = RI_h + r_1 x_1 + r_2 x_1^2 + \cdots + r_{k-1} x_1^{k-1} \pmod p \\ rw_{h,2} = RI_h + r_1 x_2 + r_2 x_2^2 + \cdots + r_{k-1} x_2^{k-1} \pmod p \\ \vdots \\ rw_{h,n} = RI_h + r_1 x_n + r_2 x_n^2 + \cdots + r_{k-1} x_n^{k-1} \pmod p \end{cases}$$

となる．

$$\begin{cases} q_1 = r_1 x_1 + r_2 x_1^2 + \cdots + r_{k-1} x_1^{k-1} \\ q_2 = r_1 x_2 + r_2 x_2^2 + \cdots + r_{k-1} x_2^{k-1} \\ \vdots \\ q_n = r_1 x_n + r_2 x_n^2 + \cdots + r_{k-1} x_n^{k-1} \end{cases}$$

とおくと，

$$\begin{cases} rw_{g,1} = RI_g + q_1 \pmod p \\ rw_{g,2} = RI_g + q_2 \pmod p \\ \vdots \\ rw_{g,n} = RI_g + q_n \pmod p \end{cases}$$

$$\begin{cases} rw_{h,1} = RI_h + q_1 \pmod p \\ rw_{h,2} = RI_h + q_2 \pmod p \\ \vdots \\ rw_{h,n} = RI_h + q_n \pmod p \end{cases}$$

とすることができる．つまり，

$$rw_{g,i} = RI_g + q_i \pmod p$$
$$rw_{h,i} = RI_h + q_i \pmod p$$

とすることができる．

$RI_g \neq RI_h$ のとき，$(rw_{g,1}, rw_{g,2}, \cdots, rw_{g,n}) = (rw_{h,1}, rw_{h,2}, \cdots, rw_{h,n})$ であると仮定する．

$$RI_g + q_i \equiv RI_h + q_i \pmod p$$

となり，

$$RI_g \equiv RI_h \pmod p$$

とすることができる．$RI_g, RI_h < p$ であるため

$$RI_g = RI_h$$

であることが分かる．すべての i についても同様に言える．つまり，仮定に矛盾し，$RI_g \neq RI_h$ のとき，$(rw_{g,1}, rw_{g,2}, \cdots, rw_{g,n}) \neq (rw_{h,1}, rw_{h,2}, \cdots, rw_{h,n})$ である．検索キーシェアについても同様のことが言える．したがって，検索用シェアと検索キーシェアはそれぞれ一意である．

3.3　秘密分散データを用いた復元不要な検索

本節では，部分復元可能な秘密分散法における秘密分散データを用いた検索について述べる．提案する検索方法は，検索用の情報を付与する分散段階と，検索段階の 2 段階からなる．

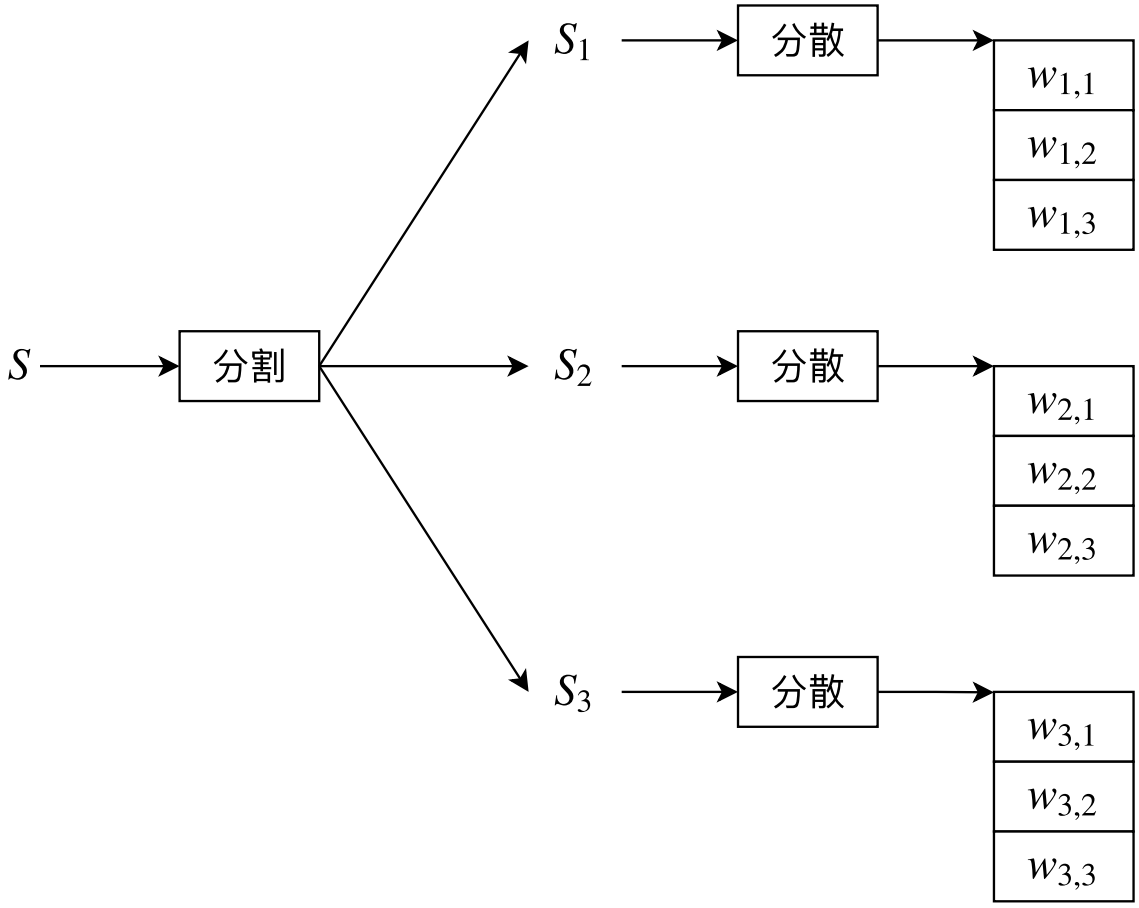

図 3.1　秘密分散データを用いた検索の医療データの分散の流れ

3.3.1　分散段階

分散したいデータを S とする．S の分散の流れを図 3.1 検索用の情報の分散の流れを図 3.2，検索用の情報の付与の流れを図 3.3 に示す． S を部分復元可能な秘密分散してシェアを生成する．S の姓情報を利用し，検索用の情報 RI を生成する．RI を一定のデータサイズ d で 3 個に分けたものを $RI_i(i = 1, 2, 3)$ とする．一定のデータサイズ d で 3 個に分けることを，分割とする．一定の素数を $p(RI_i < p$ かつ $n < p)$，$\mathbb{Z}/p\mathbb{Z} - \{0\}$ 上の集合を X，$\mathbb{Z}/p\mathbb{Z}$ 上の集合を R とおく．すなわち X は

$$X = \{x_1, x_2, \cdots, x_n\}$$

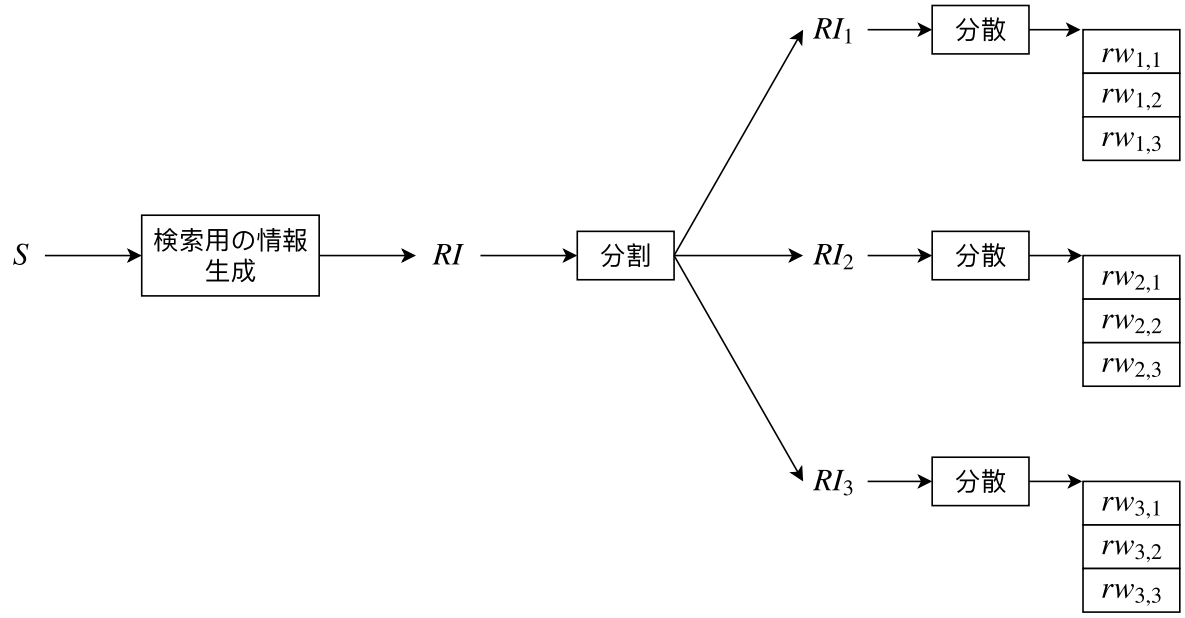

図 3.2　秘密分散データを用いた検索の検索用情報の分散の流れ

とおくことができる．X 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 vandermonde 行列 $\mathbf{X}$ を作成する．RI_i と R から $k \times 1$ の行列 $\mathbf{RA}_i$

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

を作成する．p を法とした行列 $\mathbf{X}$ と $\mathbf{RA}_i$ の乗算から

$$\mathbf{XRA}_i = \begin{pmatrix} rw_{i,1} \\ rw_{i,2} \\ \vdots \\ rw_{i,n} \end{pmatrix} \pmod{p}$$

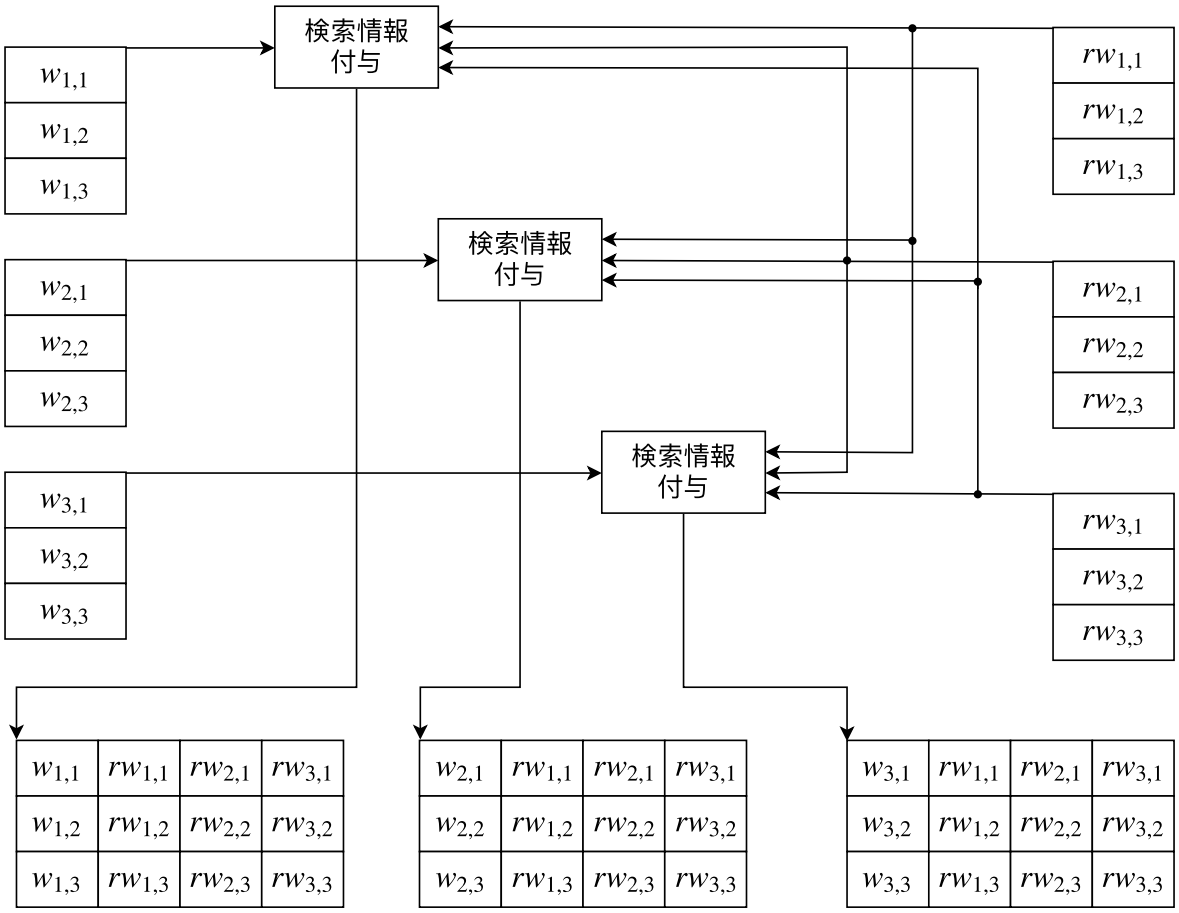

図 3.3　秘密分散データを用いた検索の検索用情報の付与の流れ

となるような検索用シェア $rw_{i,j}(j = 1, 2, \cdots, n)$ が得られる．すべての i についても同様に検索用シェアの生成を行い，S のシェアに付与する．

3.3.2　検索段階

検索キーワードの分散の流れを図 3.4，検索の流れを図 3.5 に示す．検索キーワードを RK とし，一定のデータサイズ d で 3 個に分割したものを $RK_i(i = 1, 2, 3)$ とする．分散

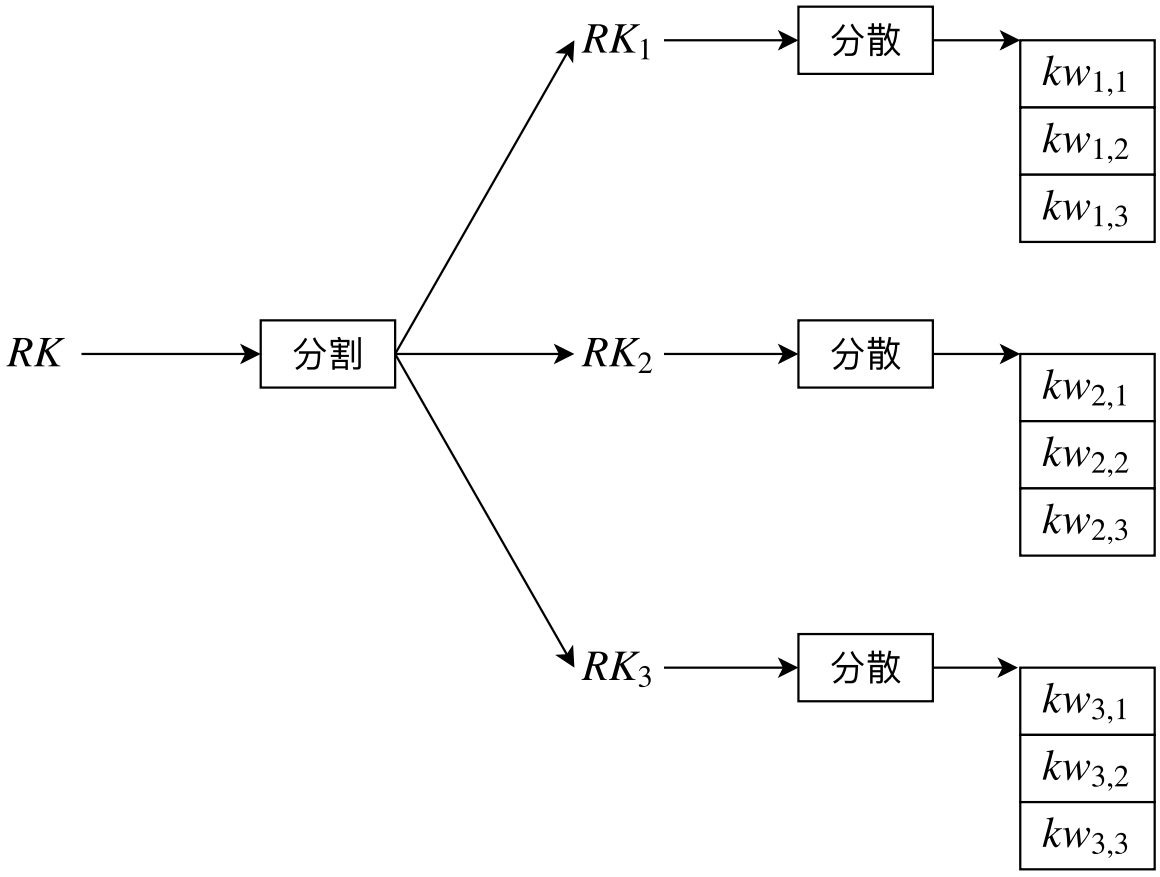

図 3.4　秘密分散データを用いた検索の検索キーワードの分散の流れ

段階で用いた集合 X，R と RK_i より，

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

となるような vandermonde 行列 $\mathbf{X}$ と，$k \times 1$ の行列 $\mathbf{KA}_i$

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

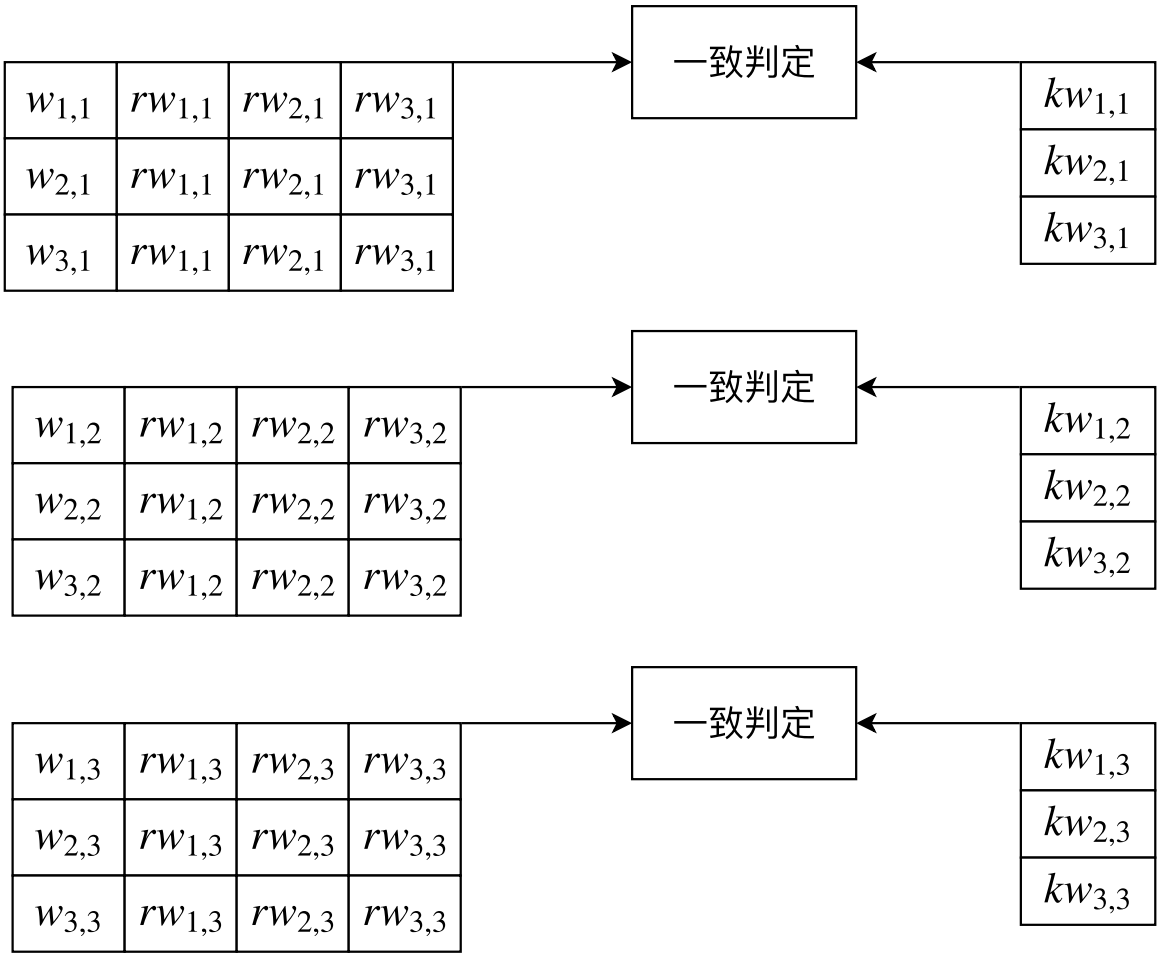

図 3.5　秘密分散データを用いた検索の検索の流れ

を作成する．p を法とした行列 $\mathbf{X}$ と $\mathbf{KA}_i$ の乗算から

$$\mathbf{XKA}_i = \begin{pmatrix} kw_{i,1} \\ kw_{i,2} \\ \vdots \\ kw_{i,n} \end{pmatrix} \pmod{p}$$

となるような検索キーシェア $kw_{i,j}(j = 1, 2, \cdots, n)$ が得られる．すべての検索用シェアのうち $i = 1$ のものと $kw_{1,j}$ の一致判定を行う．検索用シェア $(i = 1)$ と $kw_{1,j}$ の一致判定で一致したデータに対して，検索用シェアのうち $i = 2$ のものと $kw_{2,j}$ の一致判定を行う．検索用シェア $(i = 2)$ と $kw_{2,j}$ の一致判定で一致したデータに対して，検索用シェアのうち

$i = 3$ のものと $kw_{3,j}$ の一致判定を行い，一致するデータを見つける．

3.4　秘密分散データを用いた検索にかかる演算量

本節では，秘密分散データを用いた検索にかかる演算量を示す．計算コストの高い乗算演算に着目して演算量を求める．秘密分散データを用いた検索にかかる乗算回数を検索手順に沿って示す．

検索キーワード RK を一定のデータサイズ d で 3 個に分割し，$RK_i(i = 1, 2, 3)$ を得る．素数を $p(RK_i < p$ かつ $n < p)$ とする．$\mathbb{Z}/p\mathbb{Z} - \{0\}$ 上の集合 X より，$n \times k$ の vandermonde 行列 $\mathbf{X}$ を作成する．

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

vandermonde 行列 $\mathbf{X}$ の作成に必要な乗算回数 V は,

$$V = \begin{cases} 0 & (k < 3) \\ n\sum_{c=1}^{k-2} c & (otherwise) \end{cases}$$

回となる．RK_i と $\mathbb{Z}/p\mathbb{Z}$ 上の集合 R より，$k \times 1$ の行列 $\mathbf{KA}_i$

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

を作成する．行列 $\mathbf{X}$ と $\mathbf{KA}_i$ の乗算より,

$$\mathbf{XKA}_i = \begin{pmatrix} kw_{i,1} \\ kw_{i,2} \\ \vdots \\ kw_{i,n} \end{pmatrix} \pmod{p}$$

となるような検索キーシェア $kw_{i,j}(j = 1, 2, \cdots, n)$ を得る．検索キーシェア $kw_{i,j}$ の作成に必要な乗算回数 M は，

$$M = n \times k$$

回となる．以上より，分割した検索キーワード 1 個の分散に必要な乗算回数は $V + M$ となり，

$$\begin{cases} nk & (k < 3) \\ nk + n\sum_{c=1}^{k-2} c & (otherwise) \end{cases} \tag{3.5}$$

回となる．以上の操作をすべての i についても同様に行い，検索用シェアとの一致判定を行う．したがって，すべての検索キーシェアの作成に必要な乗算回数は，(3.5) 式より，

$$\begin{cases} 3nk & (k < 3) \\ 3(nk + n\sum_{c=1}^{k-2} c) & (otherwise) \end{cases} \tag{3.6}$$

回となる．1 回の乗算にかかる乗算の演算量を d^2 とすると，(3.6) 式より，検索に必要な演算量は

$$\begin{cases} 3d^2nk & (k < 3) \\ 3d^2(nk + n\sum_{c=1}^{k-2} c) & (otherwise) \end{cases} \tag{3.7}$$

となる．

3.5　従来検索方法と提案検索方法の比較

本節では，従来の検索方法と提案検索方法の検索にかかる演算量と安全性について比較する．

3.5.1　検索にかかる演算量の比較

表 2.3 の 1 文字ずつ分散して 1 文字ずつ復元し検索を行う場合の従来検索方法の演算量と，(3.7) 式より提案検索方法の演算量を求め，比較する．平仮名 1 文字のデータサイズを

表 3.1　従来の検索方法と提案検索方法の検索にかかる演算量

	従来検索方法	提案検索方法
演算量	9,231,146,496	15,360

2bytes とし，秘密分散法のしきい値を $(3,5)$ しきい値とする．従来の検索方法と提案検索方法の検索にかかる演算量を表 3.1 に示す．表 3.1 より，演算量を削減できていることが分かる．また，従来の検索方法は医療データのシェアの数に比例して演算量が大きくなるため，医療データのシェアの数が増加すると，検索にかかる演算量の差がより広がる．

3.5.2　安全性の比較

従来の検索方法では，検索用の情報の分散には乱数を用いている．よって，ある検索用の情報の分散に用いた乱数１組が漏洩しても，全ての検索用の情報を知られてしまうことは無い．しかし，提案検索方法では，検索用の情報の分散には定数を用いている．したがって，定数が漏洩した場合，全ての検索用の情報を知られてしまう．また，検索用シェアを見ると，同じ検索用の情報が付与されているデータが分かってしまう．以上より，提案検索方法は従来の検索方法より安全性が劣っていることがわかる．今後は，提案検索方法の安全性を確保する方法の考案が必要である．

第 4 章

医療データ検索方法と秘密分散データを用いた検索の実装

部分復元可能な秘密分散バックアップした医療データ検索方法と，部分復元可能な秘密分散法における秘密分散データを用いた検索の，検索にかかる時間の比較を行うことを目的とし実装した．本章では，実装した 2 つの検索法の設計について述べる．

4.1 医療データ検索方法

本節では，部分復元可能な秘密分散バックアップした医療データ検索方法の設計について述べる．医療データ検索方法は，検索情報を付与する分散段階と，検索段階に分けて設計した．(k, n) しきい値，検索用の情報の分割データサイズ d，素数 p などのパラメータはプログラムに記述しておき，分散したいデータ S はファイルに記述しておく．分散段階の処理の流れを図 4.1 に示し，検索段階の処理の流れを図 4.2 に示す．

4.2 秘密分散データを用いた検索

本節では，部分復元可能な秘密分散法における秘密分散データを用いた検索の設計について述べる．秘密分散データを用いた検索は，検索情報を付与する分散段階と，検索段階に分けて設計した．(k, n) しきい値，検索用の情報の分割データサイズ d，素数 p，集合 X，R，などのパラメータはプログラムに記述しておき，分散したいデータ S はファイルに記述しておく．分散段階の処理の流れを図 4.3 に示し，検索段階の処理の流れを図 4.4 に示す．

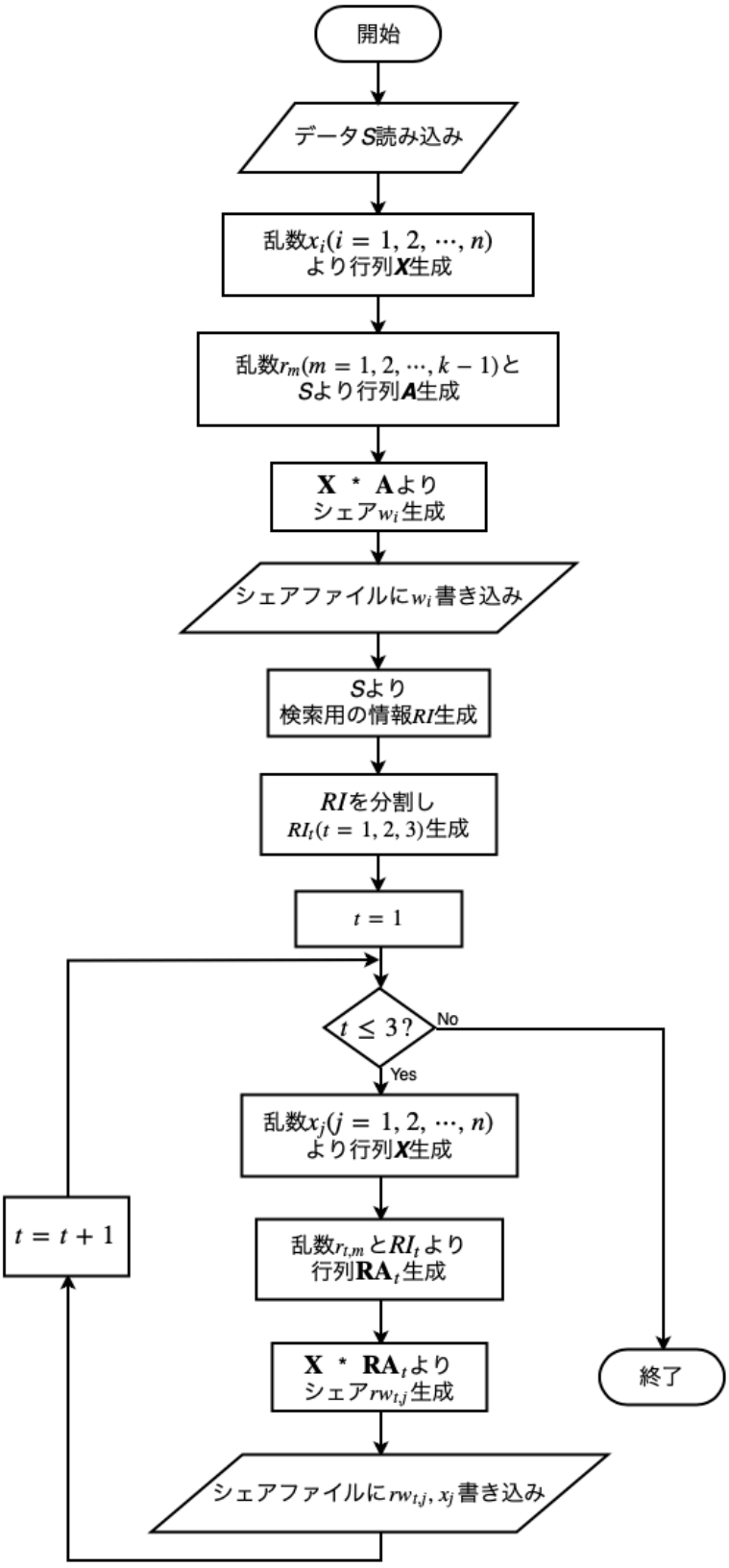

図 4.1　医療データ検索方法の分散段階の処理の流れ

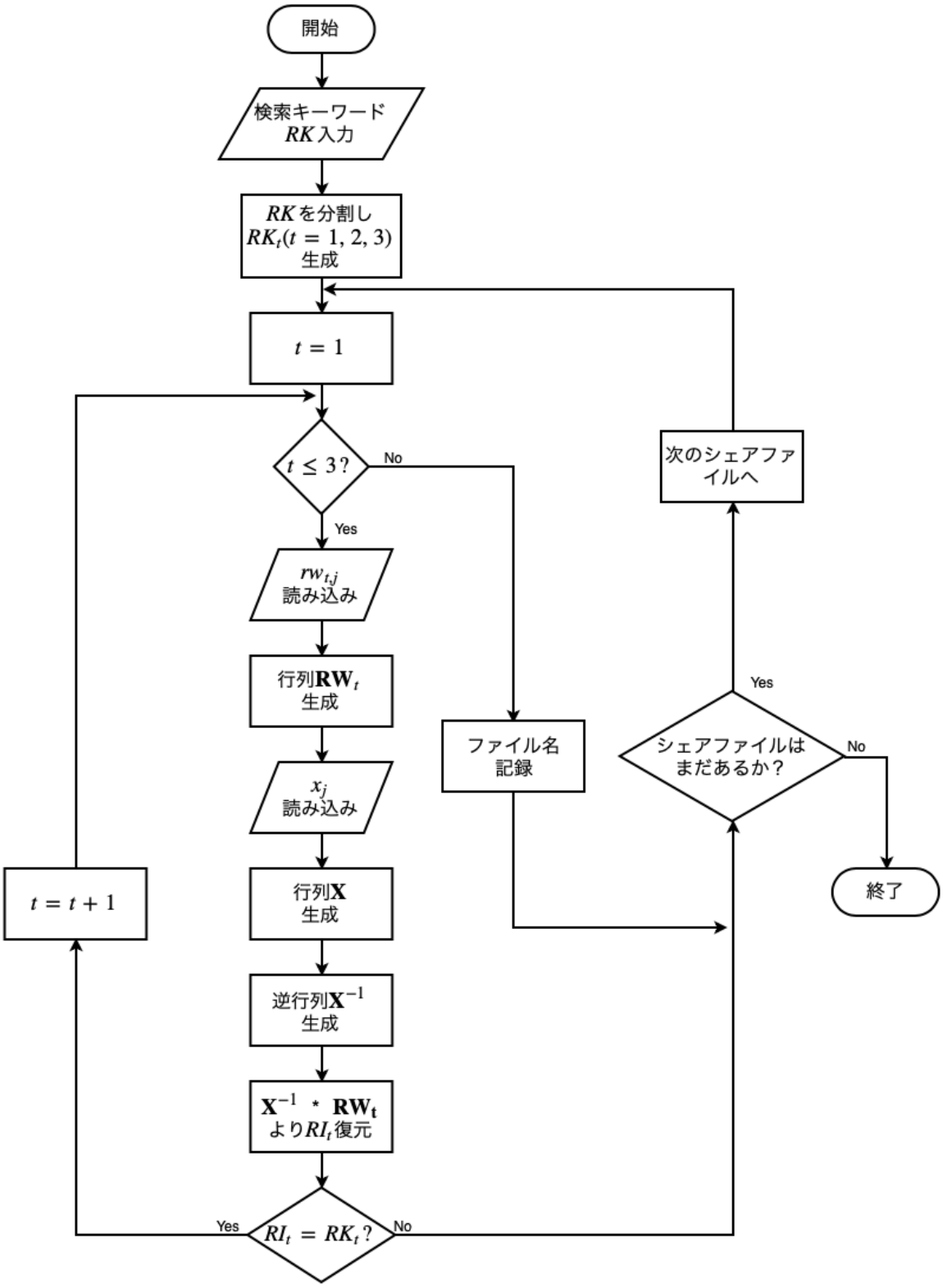

図 4.2　医療データ検索方法の検索段階の処理の流れ

4.2　秘密分散データを用いた検索

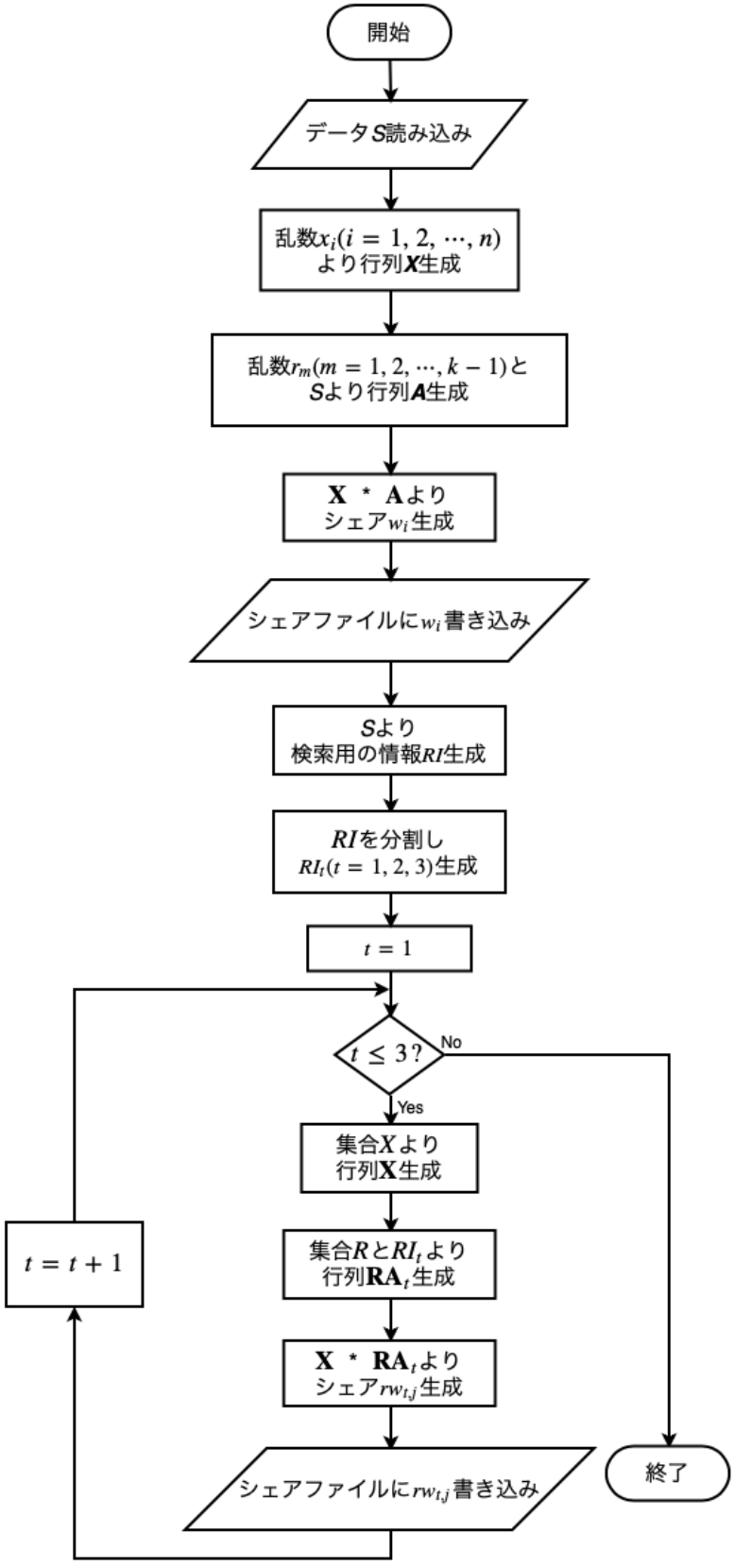

図 4.3　秘密分散データを用いた検索の分散段階の処理の流れ

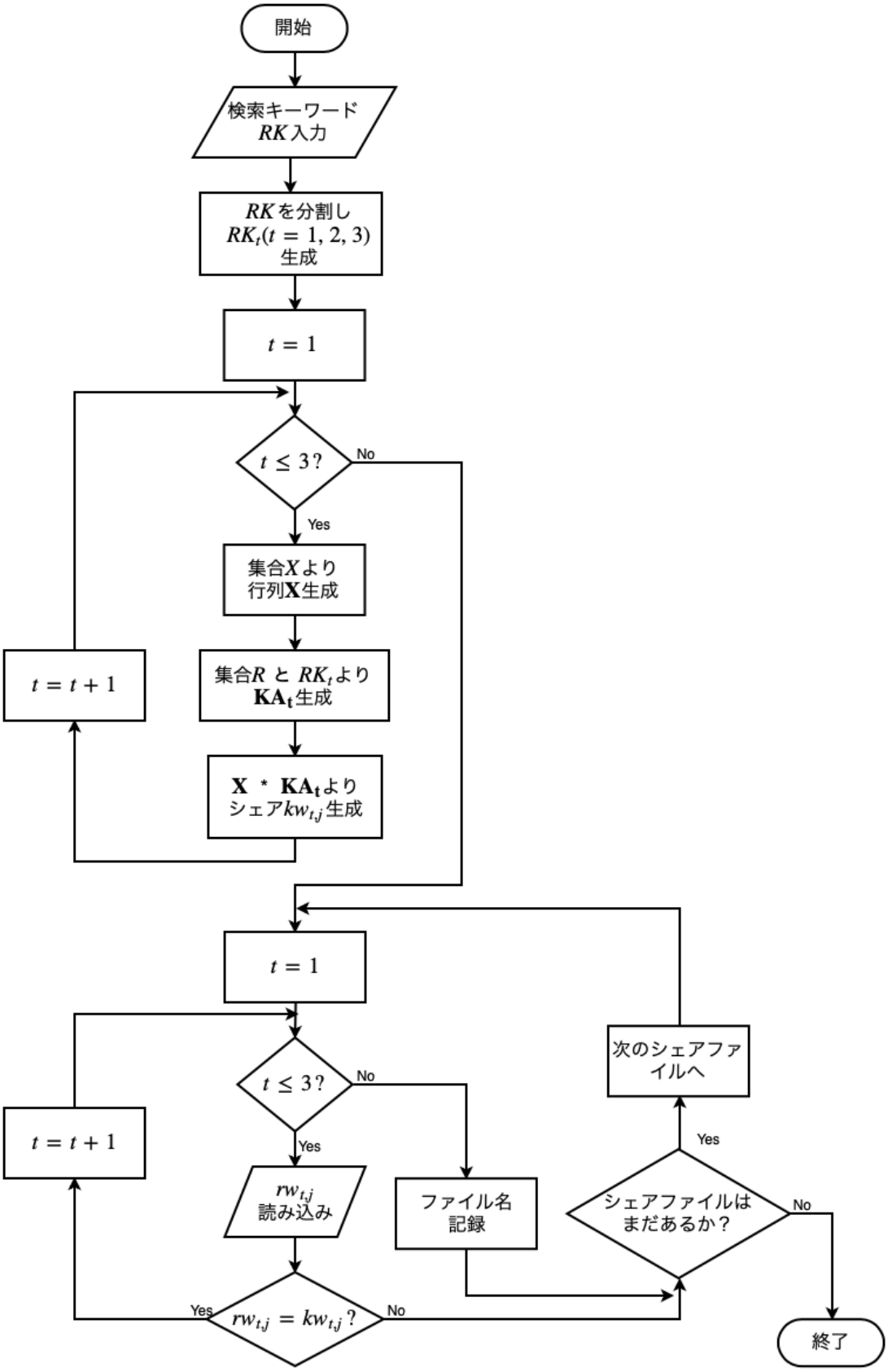

図 4.4　秘密分散データを用いた検索の検索段階の処理の流れ

第 5 章

検索にかかる時間の比較

本章では，本研究で実装した部分復元可能な秘密分散バックアップした医療データ検索方法と，部分復元可能な秘密分散法における秘密分散データを用いた検索の検索にかかる時間を比較する．

5.1 比較実験

データの総数は 100000 個, 200000 個, $\cdots$, 1000000 個 であり，$(3,5)$ しきい値秘密分散法を用いているため，データのシェアの総数は 100000×5 個, 200000×5 個, $\cdots$, 1000000×5 個のうち，検索で一致するデータが 20000×5 個である場合の検索にかかった時間を計測した．また，6bytes の検索用の情報を 2bytes ずつ分割して $(3,5)$ しきい値秘密分散法で分散し，データのシェアに付与した．計測した時間は，従来の検索方法では 4.1 節検索段階の処理にかかった時間，秘密分散データを用いた検索では，4.2 節検索段階の処理にかかった時間である．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．

5.1.1 実験環境

実験に用いた環境を表 5.1 に示す．

表 5.1　実験環境

OS	Ubuntu18.04.2 LTS
CPU	Intel ®Xeon E3-1230 v6 3.5GHz
メモリ	16GB
コンパイラ	g++-7.3
コンパイラオプション	-std=c++11

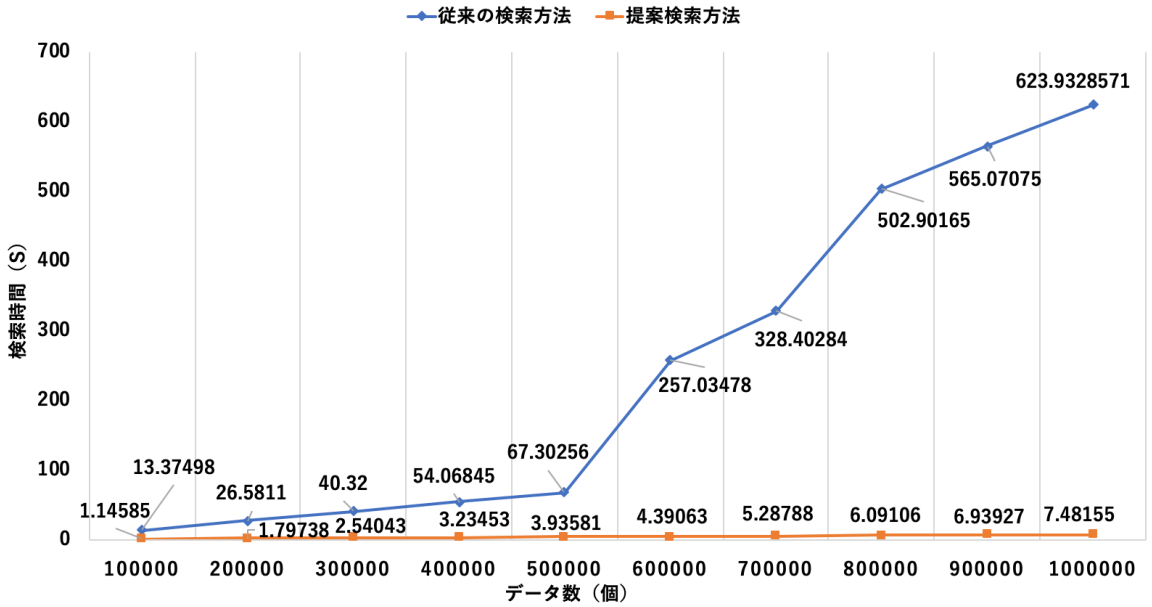

図 5.1　従来の検索方法と提案検索方法の検索にかかった時間

5.1.2　結果

従来の検索方法と，秘密分散データを用いた検索の検索にかかった時間を図 5.1 に示す．縦軸は検索にかかった時間であり，横軸はデータの総数を表している．図 5.1 より，秘密分散データを用いた検索は，従来の検索方法より検索にかかる時間を短縮できていることが分かる．また，データ数が増加するごとに検索にかかる時間の差が広がることが確認できる．

5.1.3　考察

図 5.1 より，検索にかかる時間を短縮できていることが確認できた．高知県の人口約 70 万人を考え，医療データが 1 人 1 個だと仮定すると，提案検索方法の検索にかかる時間は約 5 秒と十分高速であると考える [6]．また，医療データがそれ以上である場合も，十分高速に検索できると考えるが，提案検索方法は線形探索で検索を行っているため，二分探索のような探索法を用いることで，より高速に検索できると考える．

第 6 章

結論

6.1 本研究のまとめ

本研究では，部分復元可能な秘密分散システムにおける検索の高速化を目的として，秘密分散データ同士での一致の判定を可能とする方法を提案し，提案方法で検索ができることの証明を行った．そして，医療データ検索方法と秘密分散データを用いた検索の実装と検索にかかる時間の比較を行った．

秘密分散データを用いた検索では，すべての検索用の情報を定数を用いて (k, n) しきい値秘密分散法で分散して医療データに付与し，検索時は分散段階で用いた定数を用いて検索キーワードを (k, n) しきい値秘密分散法で分散した．この結果，検索用の情報と検索キーワードをシェアの状態で一致判定することが可能となり，検索用の情報の復元処理を省略できるため，検索にかかる演算量を小さくすることが出来た．しかし，定数を用いたことにより，従来の検索方法より安全性が劣っていたため，安全性を確保する方法を考案する必要がある．

部分復元可能な秘密分散バックアップした医療データ検索方法と秘密分散データを用いた検索の検索にかかる時間の比較を行った結果，検索にかかる時間を短縮できていることが確認できた．また，検索用シェアの探索法を変えることで，検索にかかる時間のさらなる短縮が期待できる．

6.2　今後の課題

今後の課題として，提案検索法の安全性を確保する方法については考えられていないため，考える必要がある．また，提案検索方法では，線形探索で検索を行っているため，検索をより高速化するために，シェアファイルの探索法を考える必要がある．

謝辞

本研究を行うにあたり，ご指導頂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．理解した風を装ってなんとなくこなしていこうという私を見抜き，呆れながらも何度もご指導してくださったこと大変感謝しています．本研究の副査をしていただいた情報学群敷田幹文教授，鵜川始陽准教授のお二人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．

NOC の職員であり，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．何度もお食事に連れて行ってもらったり，研究についてのアドバイスを頂きました．時には一緒に厨房に立って料理を作ったり？大変楽しい時間を過ごすことができました．

Bandhit.Suksiri 氏にも謹んで感謝致します．研究についてのアドバイスや，日常生活では使わないであろう各国の言葉を教えて頂きました．一緒に作ったガイヤーン，トムヤムクン，ガパオライスの味は一生忘れません．

同期のぽよんこと横江良哉氏にも謹んで感謝致します．ぽよん氏には，持ち前の「まあええんちゃう？」精神で，何度も支えられました．最後は 2 人になっていたりと，寂しさを感じたときもありましたが，ぽよん氏のおかげで楽しかったです．

福本研究室 3 年生の皆さん，たまに研究の息抜きに付き合って頂きありがとうございました．よく仕事を押し付けていた小野田祐稀氏には大変感謝しています．

最後になりましたが，4 年間の大学生活を支えてくださった皆様に感謝致します．

参考文献

[1] 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.

[2] 沼尚樹，“部分復元可能な秘密分散バックアップした医療データ検索方法,”平成 29 年度高知工科大学学士学位論文，Mar. 2018.

[3] 福本昌弘，“南海トラフ巨大地震に対する医療情報の保全のための高知県での取り組み,”Mercato，東北情報通信懇談会，Vol. 89，pp. 18–20，2014.

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

[5] 田中康仁，“日本人の姓と名の分布,”http://www.orsj.or.jp/~archive/pdf/bul/Vol.23_06_357.pdf，日本オペレーションズ・リサーチ学会，p. 361，Jun. 1978.

[6] 高知県総務部統計分析課，“高知県の推計人口年報 (平成 30 年)～平成 30 年 10 月 1 日現在～,”http://www.pref.kochi.lg.jp/soshiki/111901/files/2014021401751/h30nenpou.pdf，2019 年 2 月 25 日閲覧.