平成 25 年度委託研究

「統計データの提供における情報セキュリティ技術の応用に関する実証研究」報告書

平成 26 年 3 月 14 日

NTT セキュアプラットフォーム研究所

目次

1 はじめに

1.1 目的

本研究は、現在 NTT セキュアプラットフォーム研究所で研究が進められている秘密分散・秘密計算技術を用いて、汎用集計ソフトウェアと連係した、リモートアクセス型のオーダーメード集計の可能性と課題について総務省統計研修所の平成25年度委託研究として実証研究を行う事を目的とする。

1.2 背景

現在、総務省統計局では統計法に基づき、学術研究の発展や、高等教育の発展に資する事を目的として、「匿名データの作成・提供」、「オーダーメード集計」サービスを提供している。サービスの運用は独立行政法人 統計センターが行っており、以下のような利用者が想定されている。

- 大学等や学術研究を目的とする機関に所属する研究者又は当該機関
- シンクタンク等で学術研究を行う者又は当該機関
- 機関に所属していないが、学術研究を行っている研究者
- 大学等の高等教育機関において講義等の教育を行う教員又は当該機関

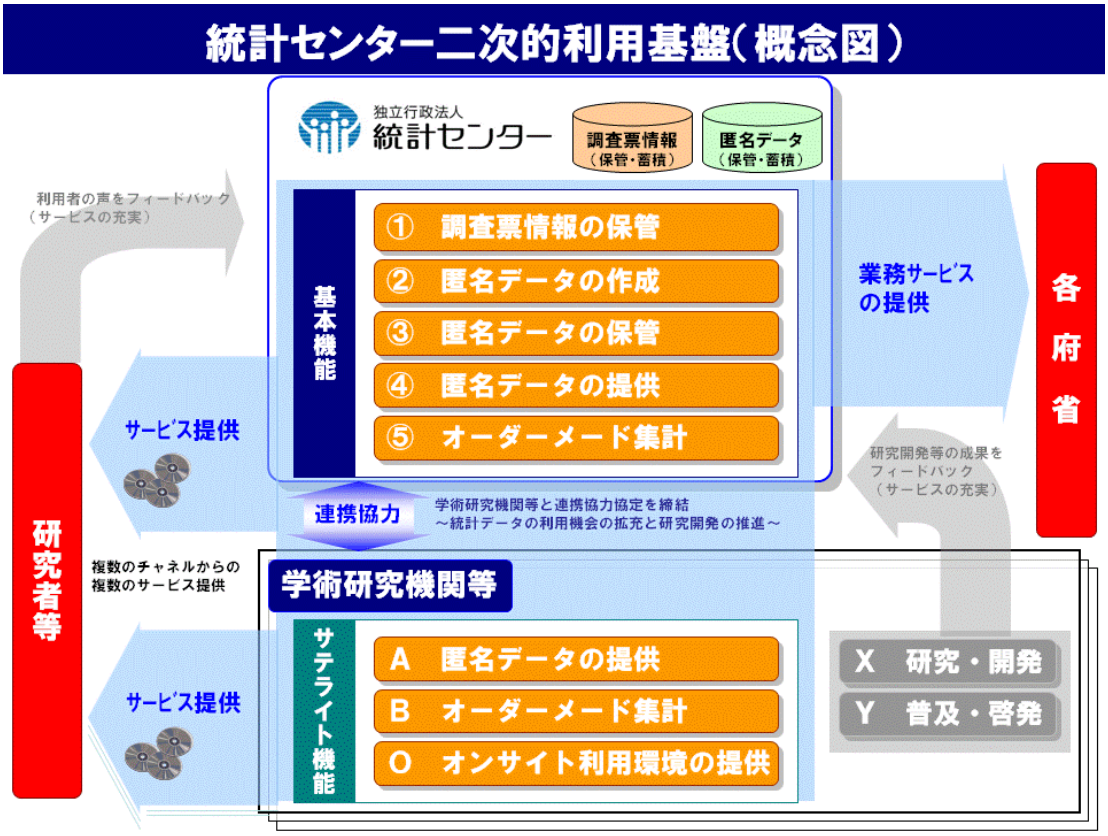


図 1.1　統計センター2 次的利用基盤（概念図）

出典：統計センターウェブページ http://www.nstac.go.jp/services/archives.html

本研究では主に「オーダーメード集計」を対象として実証研究を行い、「匿名データの作成・提供」については課題としてあげるに留める。オーダーメード集計とは総務省統計局が委託に応じて、統計調査から集められた情報を利用して統計等を作成し提供するものである。現在対象となる統計表は下表の通りである。

表 1.1　対象となる統計表

対象となる 統計調査名	調査の年次
国勢調査	昭和 55 年、昭和 60 年、平成 2 年、平成 7 年、平成 12 年、平成 17 年
労働力調査	昭和 55 年 1 月～平成 24 年 12 月
家計消費状況調査	平成 14 年 1 月～平成 24 年 12 月

住宅・土地統計調査	昭和 63 年、平成 5 年、平成 10 年、平成 15 年、平成 20 年
就業構造基本調査	昭和 57 年、昭和 62 年、平成 4 年、平成 9 年、平成 14 年、平成 19 年
社会生活基本調査	昭和 56 年、昭和 61 年、平成 3 年、平成 8 年、平成 13 年、平成 18 年、平成 23 年
全国消費実態調査	平成 16 年、平成 21 年
家計調査	平成元年 1 月～平成 24 年 12 月

これらのサービスを利用する際には、事前に定められた手続きを行い、後日申請結果を受け取るか、指定されたサテライト機関内で利用する事になる。

図 1.2　現在のオーダーメード集計の利用方法

出典：統計センターウェブページ http://www.nstac.go.jp/services/order.html

オーストラリアやニュージランドなどの諸外国では既にリモートアクセス型のオーダーメード集計等のサービスが開始されており、今後は国内でも厳正な情報セキュリティを維持した状態で、より利便性の高いリモートアクセス型のオーダーメード集計の実現が求められる。

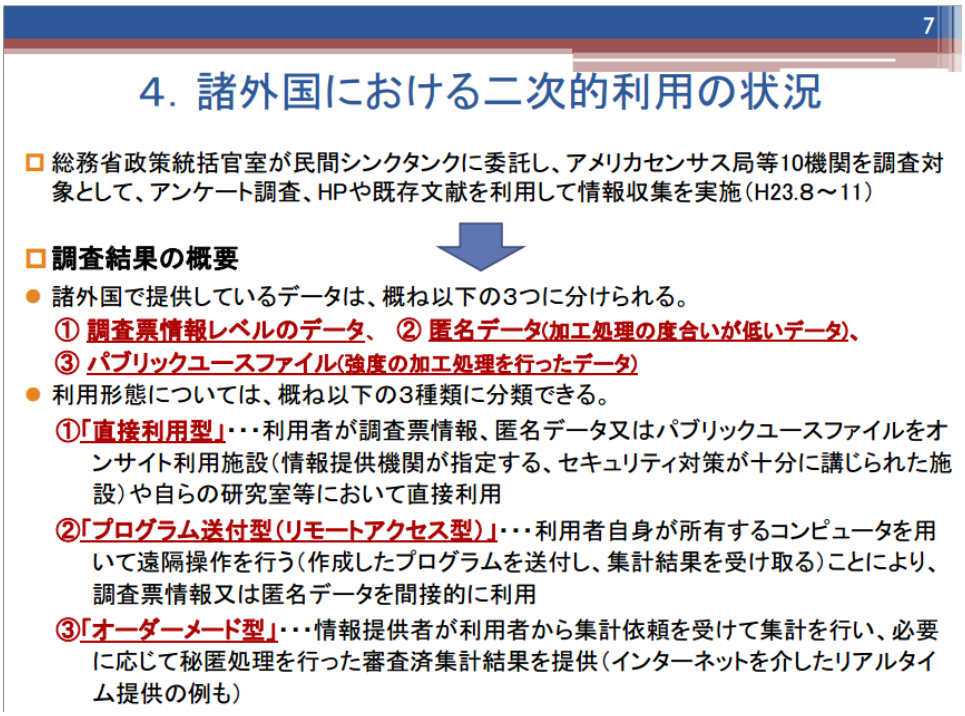


図 1.3 諸外国における 2 次的利用の状況

出典：総務省ウェブページ http://www.nstac.go.jp/services/pdf/121116_5-1.pdf

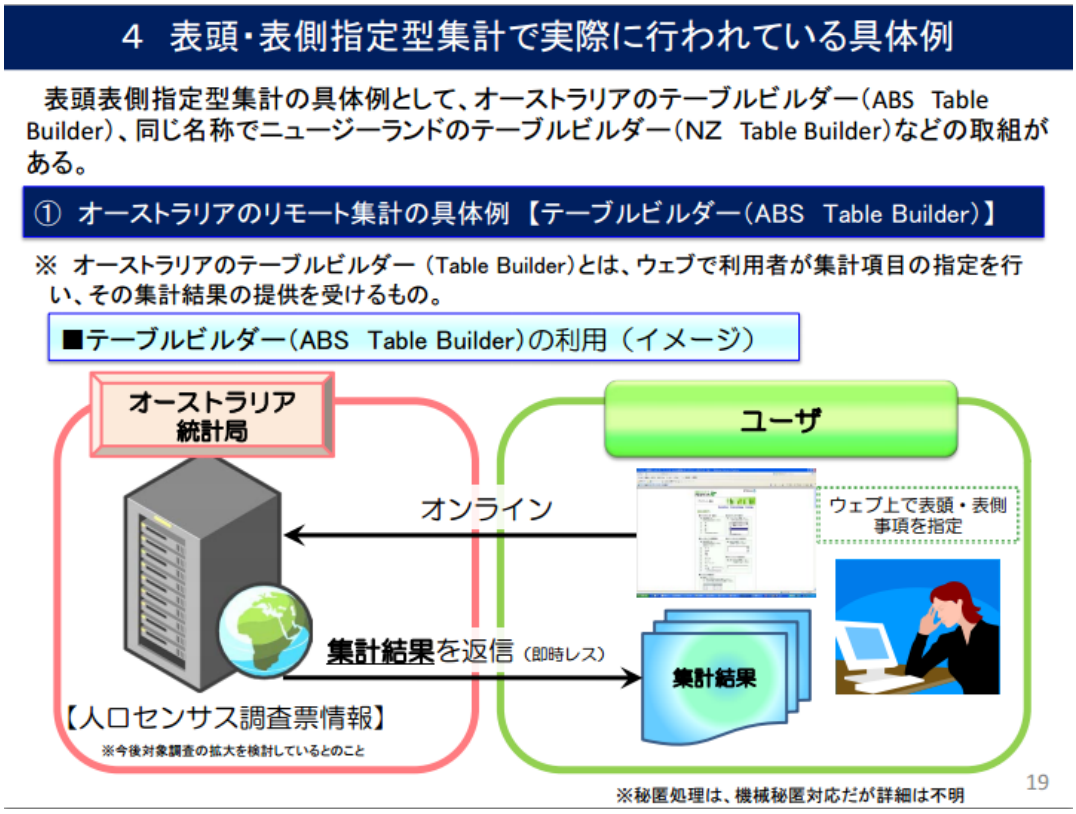


図 1.4 オーストラリアでのリモートアクセス型集計の例

出典：総務省ウェブページ http://www.stat.go.jp/info/kenkyu/2jiriyou/pdf/si11-4.pdf

その為、今回は秘密分散・秘密計算技術を用いて、強度なセキュリティを維持した状態でのリモートアクセス型オーダーメード集計の可能性の調査・研究、及び実現に向けた課題の洗い出しを行う。

1.3 実証研究概要

本研究では、リモートアクセス型のオーダーメード集計の実現性の確認と課題の洗い出しの為、以下の内容を実施する。

- 評価対象製品の調査と選定
- 評価対象製品への組み込みと集計の実証
- 実証結果の評価と課題の確認

1.4 用語の定義

表 1.2　用語の定義

用語	意味
秘密計算	データを暗号化したまま処理し、処理を実行する側にはデータを一切秘密にしたまま、処理結果のみを得る事ができる技術。
秘密分散	データを暗号化し、複数のサーバに分散配置する技術。
公的データ	主に中央官庁が公開している統計表や、その基礎となるデータそのものを指す。
オーダーメード集計	既存の統計調査で得られた調査票データを活用して、調査実施機関等が申出者からの委託を受けて、そのオーダーに基づいた新たな統計表を集計・作成し、提供するもの。
匿名データ	行政機関等が統計法に基づいて実施した統計調査によって集められた調査票情報を、特定の個人又は法人その他の団体の識別（他の情報との照合による識別を含む）ができないように加工したもの。
秘匿処理	統計調査の集計結果表を作成する際、ある区分に該当する対象数が少なく、その結果数値を公表する事により、調査客体の個別の情報が判明してしまうおそれがある場合に、該当するセルを実際の数値ではなく別の値に置き換える等の処理。

2　リモートアクセス型オーダーメード集計の提供に適した表計算ソフトウェアの調査

2.1 集計表作成ソフトウェアの要件

リモートアクセス型オーダーメード集計の可能性を調査するにあたり、秘密分散・秘密計算技術を用いて集計表作成を行う汎用的集計ソフトウェアの選定を行う。ソフトウェアの選定は、以下に示す現在のオーダーメード集計の運用実態及び、今後期待される機能要件を考慮して実施する。

- Microsoft Windows 上で動作するソフトウェア
- 国内で購買できる汎用ソフトウェア
- 日本語環境
- GUI が整備され、操作が容易である
- 多重クロスが簡単に指定できる
- 連続値を簡単にカテゴライズする事ができる
- 価格は 1 ライセンス 30 万円程度までが望ましい

現在のオーダーメード集計では利用者は下図のような統計表作成仕様書を作成し、依頼を行う。利用者はどのようなデータをどのような条件で出力して欲しいのか、また表頭、表側にどのような項目を配置して欲しいのかを記入し、申請する。利用者が得る統計表は下図のような集計表であり、その為、オーダーメード集計に用いられるソフトウェアに求められる機能としては、特に「多重クロスが簡単にできる」、「連続値を簡単にカテゴライズする事ができる」の項目が重要である。

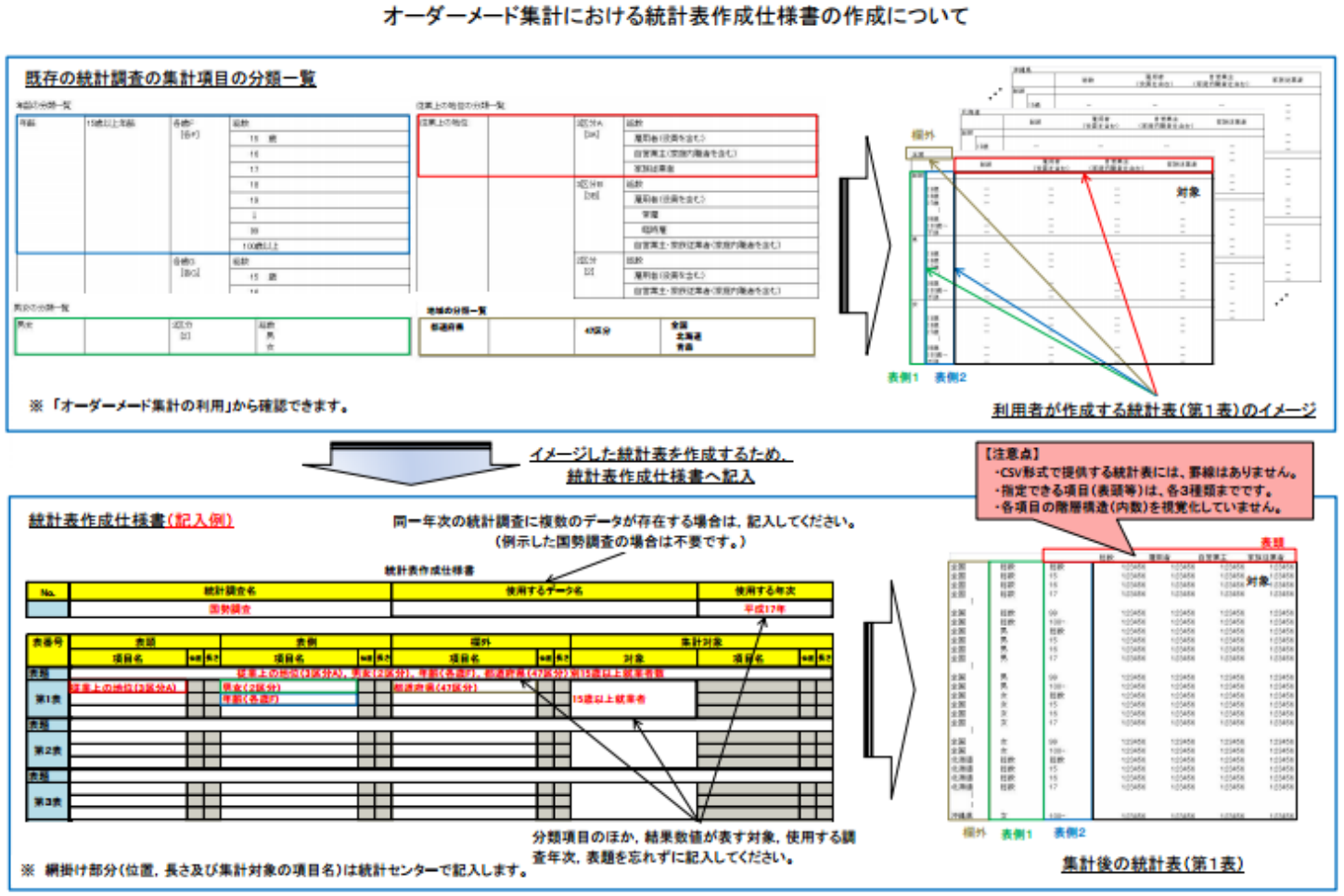


図 2.1 オーダーメード集計における統計表作成仕様書の作成について

出典：統計センターウェブページ http://www.nstac.go.jp/services/2ji/order_shiyousyo.pdf

2.2 評価対象の選定

2.1 節に示す要件から以下の 4 製品を選定し、要件に対する対応状況を評価した。

- R
- Crystal Reports
- Access
- Adam-Lite

一般にデータを扱うソフトウェアとしては、Microsoft 社の SQLServer や Oracle 社の Oracle 等の代表的な RDBMS がよく知られているが、これらは技術者向けの比較的操作の難しいシステムであり、公的データを利用するユーザーにとっては敷居が高い為、今回の調査の対象からは除いた。ただし、Access は RDBMS の一種ではあるが、一般ユーザーにも利用されているケースが多い為対象に含めることとした。

2.2.1 R

出典: http://www.r-project.org/

R は有名な統計言語である「S 言語」をオープンソースとして実装し直した統計解析ソフトウェアである。さまざまなプラットフォーム（OS）に対応しており、誰でも自由にダウンロードする事ができる。さらに、世界中の専門家が開発に携わっており、日々新しい手法・アルゴリズムが付け加えられている。

販売元	R Development Core Team
価格	無償（GNU GPL）
製品 HP	http://www.r-project.org/index.html

ただし、R 自体は主にスクリプト言語を使用し、専用のコンソール画面を用いて集計指示を行うことが想定されているソフトウェアであり、GUI 環境はあまり整備されていない。

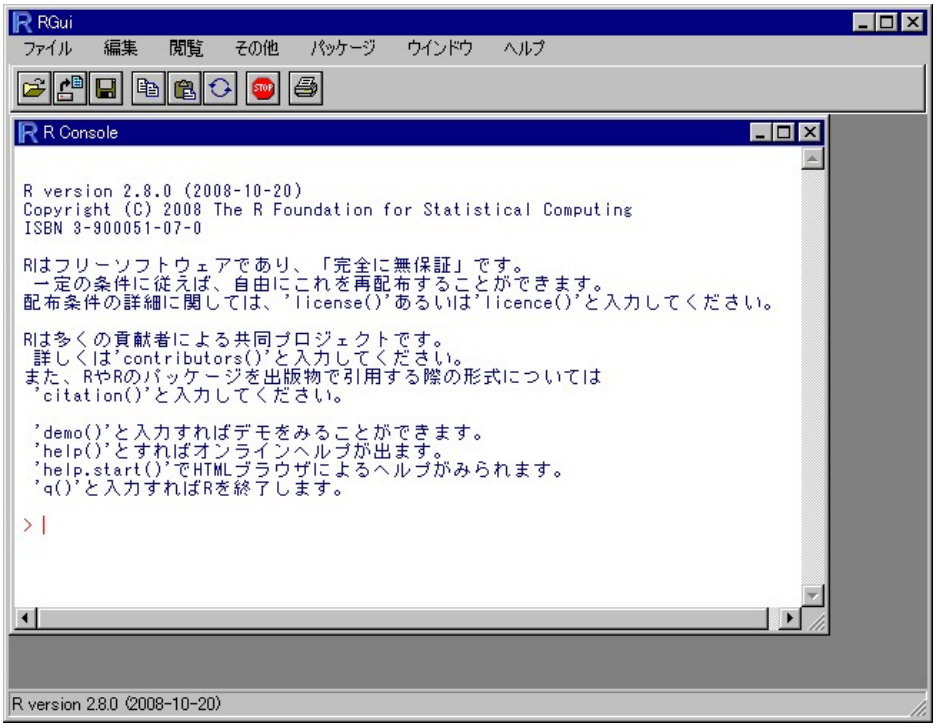


図 2.2 R のコマンドライン入力コンソール

2.2.2 SAP Crystal Reports

出典: http://crystalreports.jp/product/cs2013/index.html

Crystal Reports は、数多くのデータ・ソース型式に対応した帳票作成ソフトウェアである。Microsoft 社が提供するソフトウェア開発環境である Visual Studio にも OEM 搭載されていた事から多くの業務系システムで採用されている(Visual Studio 2010 からは標準搭載を廃止されている)。SAP 社が Business Object 社を買収した事から、現在は SAP Crystal Reports として販売されている。

販売元	SAP ジャパン株式会社
価格	¥85,300 ※SAP Crystal Reports 2013 1 指定ユーザー　製品ライセンス
製品 HP	http://crystalreports.jp/

Crystal Reports の特長としては

- あらゆるデータへのアクセス
- エンドユーザーが必要とするレポートの設計と書式設定
- ユーザーへ適切なレポートを適切なタイミングで配信する
- Web によるレポートの公開では、表示とスケジュールを高いセキュリティで提供
- ユーザーによる個別のレポート操作の強化
- レポーティングのアプリケーション及びポータルとの統合
- 既存のアプリケーション内にレポートの対話性を実現する
- 主要 Web アプリケーションサーバとプラットフォームへの統合

があり、多くのユーザーの指示を受けている。しかし、多重クロスや連続値のカテゴライズを行おうとすると、処理が難しくなってしまうという問題がある。

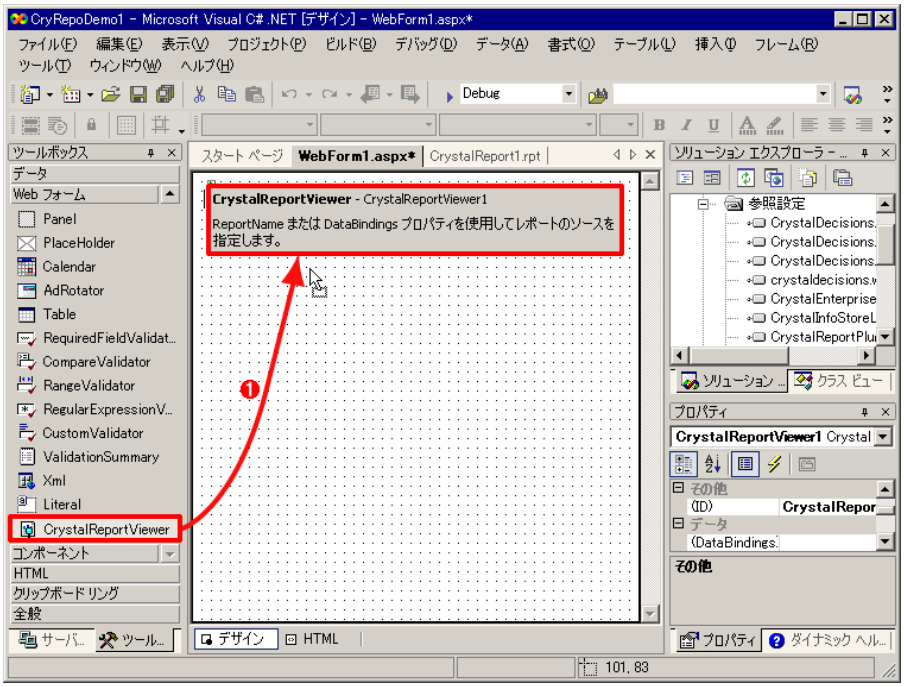


図 2.3 Crystal Reports のレポート追加画面

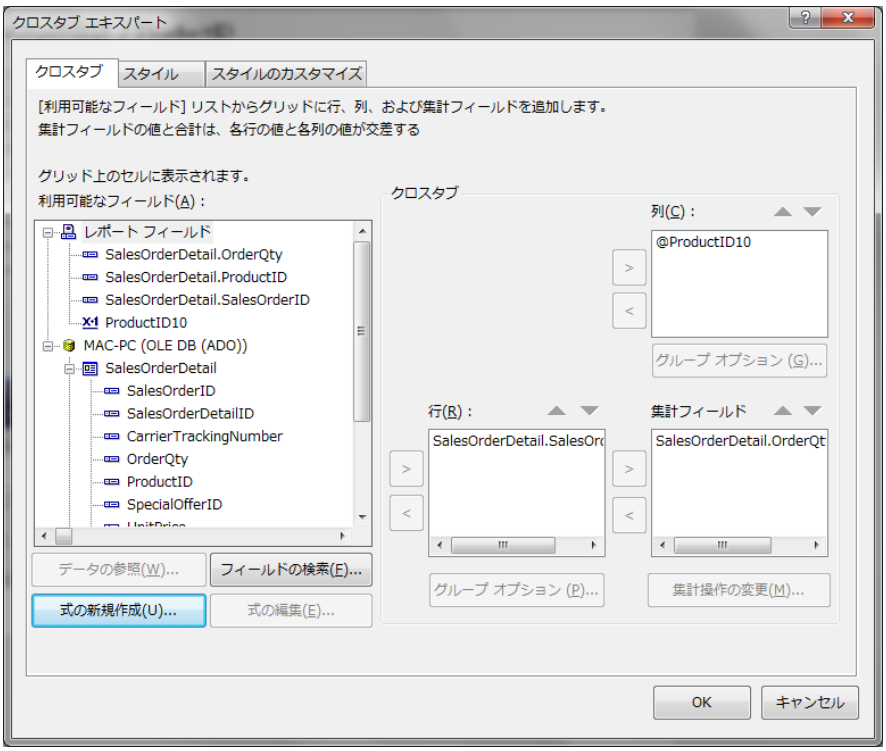


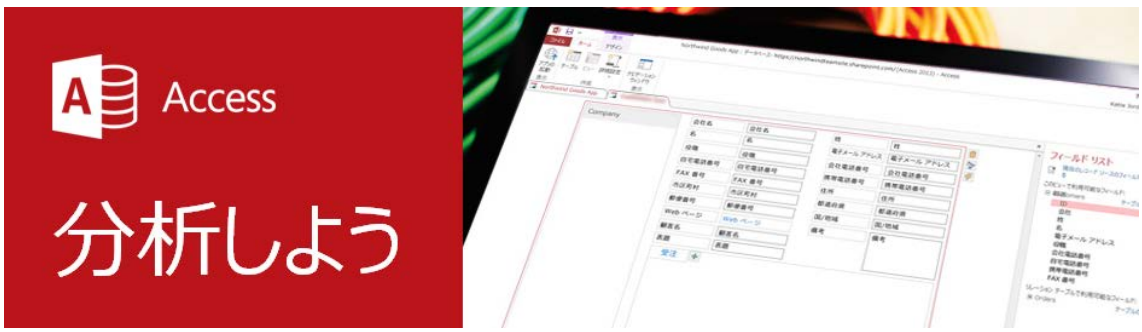
図 2.4 Crystal Reports でのクロス表作成画面

実際にクロス集計、及び連続値のカテゴライズを行う際は、まずはレポートを追加配置し、上図のようなクロス表作成画面より、項目を指定する。更に連続値のカテゴライズ用のスクリプトを記述する事になる。プログラミングに慣れた利用者には見慣れた記述ではあるが、普段プログラムを書くことのないユーザーに取っては敷居が高い。

```
Select Case {SalesOrderDetail.ProductID}
    Case 700 to 709
        formula = "700-709"
    Case 710 to 719
        formula = "710-719"
    Case 720 to 729
        formula = "720-729"
    Case 730 to 739
        formula = "730-739"
    Case 740 to 749
        formula = "740-749"
    Case 750 to 759
        formula = "750-759"
    Case 760 to 769
        formula = "760-769"
    Case 770 to 779
        formula = "770-779"
    Case 780 to 789
        formula = "780-789"
    Case 790 to 799
        formula = "790-799"
    Case Else
        formula = "irregular"
End Select
```

図 2.5 連続値のカテゴライズ記述例

2.2.3 Access



出典: http://office.microsoft.com/ja-jp/access/

Access は Microsoft Office の上位製品にバンドルされる関係データベース管理システムである。オブジェクト指向に基づいたアプリケーション開発が可能であり、帳票開発用の各種ツールも整備され、多くの業務系ソフトに使用されている。

販売元	日本マイクロソフト株式会社
価格	15,540 円
製品 HP	http://office.microsoft.com/ja-jp/access/

Access も Crystal Reports と同様にクロス集計表の作成を行う際は専用の GUI を用いて作成を行う。

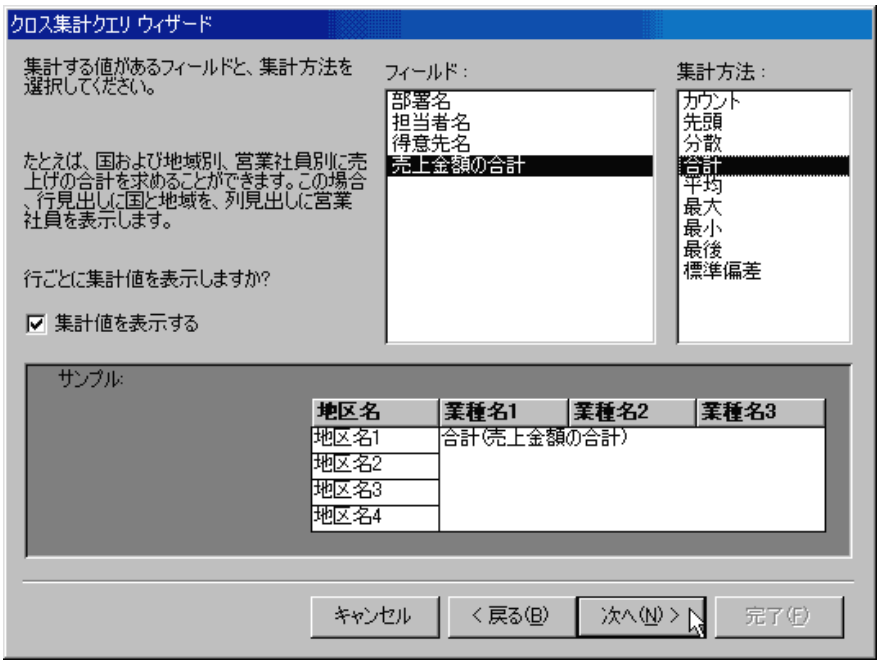
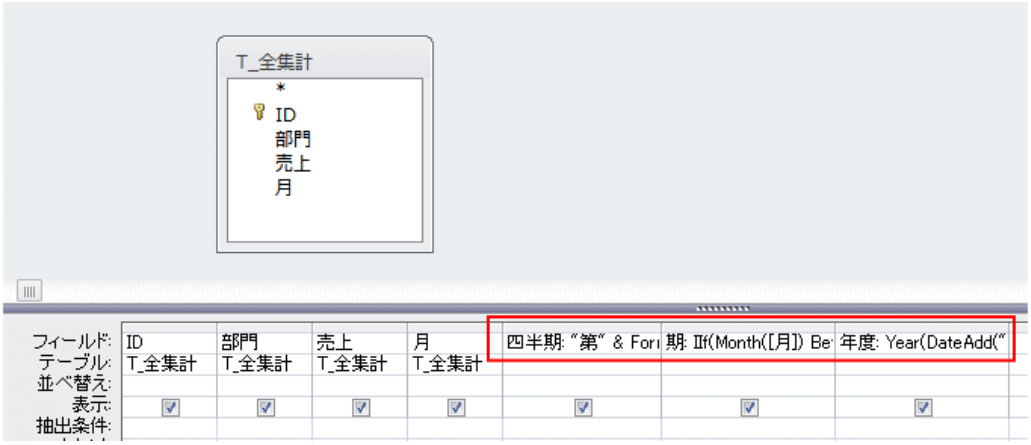


図 2.6 Access によるクロス集計表作成画面

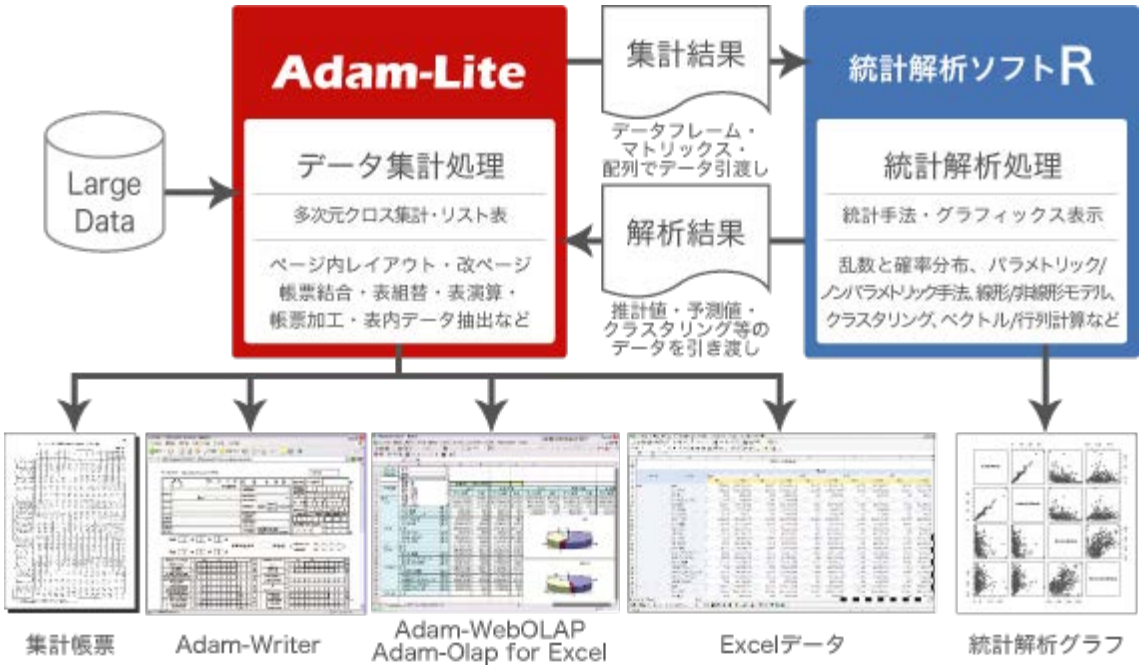



```
四半期: "第" & Format(DateAdd("m",-3,[月]),"q") & "四半期"
年度: Year(DateAdd("m",-3,[月]))
期: IIf(Month([月]) Between "4" And "9","上期","下期")
```

図 2.7 Access での連続値のカテゴライズ例

Crystal Reports と同様に連続値のカテゴライズを行う際は専用のスクリプトを記述する事になり、やはりプログラミングに不慣れなユーザーには利用が困難である。

2.2.4 Adam-Lite



出典: http://www.zetta.co.jp/products/adam_lite/

Adam-Lite は官公庁で多くの実績をもつ、Adam-Report の簡易版であり、統計解析ソフトである「R」との連係を実現したアプリケーションである。また、官公庁特有の帳票開発に必要な機能を豊富に装備しているのが特徴である。

販売元	ゼッタテクノロジー株式会社
価格	98,000 円/年
製品 HP	http://www.zetta.co.jp/products/adam_lite/

Adam-Lite を用いて多重クロス表を行う際は、下図のようなクロス表作成専用の GUI を用いて指定する事になる。表頭、表側それぞれ 3 次元まで、項目をドラッグ&ドロップするだけで簡単に作成を行う事が出来る。

図 2.8 Adam-Lite による表作成

また、連続値のカテゴライズに関しても、専用の GUI を使用して指定する事になる為、プログラミングに慣れていないユーザーにとっても、非常に簡単に利用する事が出来る。

図 2.9 Adam-Lite による連続値のカテゴライズ例

2.3 評価結果

前節で選定した各製品について、2.1 節に示す要件に対する対応状況を評価した結果を以下に示す。

表 2.1　製品別要件対応一覧

	R	Crystal Reports	Access	Adam-Lite
Microsoft Windows 上で動作するソフトウェア	○	○	○	○
国内で購買出来る汎用ソフトウェア	○	○	○	○
日本語環境	○	○	○	○
GUI が整備され、操作が容易である	×	△	△	○
多重クロスが簡単に指定できる	△	○	○	○
連続値を簡単にカテゴライズする事ができる。	△	△	△	○
価格は 1 ライセンス 30 万円程度までが望ましい	○	○	○	○

各製品とも概ね要件を満たしているが、操作性に関する項目でばらつきがある。特に連続値に対するカテゴライズを行う場合には多くの製品で専用のスクリプトを用いた記述が必要となるなど、必要とされる技術レベルが高くなる傾向にある。一方で Adam-Lite は元々官公庁内の職員が利用していた実績もあるように、この点での操作性、開発生産性は非常に高い。また、既に統計センターでの利用実績もある事から、本調査においては Adam-Lite を採用し、秘密計算技術との連係の可能性について実証研究を行う。

3 秘密分散・秘密計算技術

本節では本実証研究で用いた秘密分散および秘密計算技術の概要を説明する。

3.1 秘密分散

秘密計算技術のデータの機密性は、秘密分散の機密性に基づく。

秘密分散は秘密にしたい情報を、複数の分散情報に分けることによって保護するアルゴリズムである。分散情報を一定数集めることによって、情報を復元することができ、逆に一定数集めない限り、秘密が漏れることはない。

図 3.1 は、情報を 3 つに分散させ、2 つを集めることによって復元できる秘密分散の概念図である。このような秘密分散を実現する方法としては、Shamir の秘密分散がよく知られている。Shamir の秘密分散は多項式補間問題に基づく方法で、全体の分散数を n、復元に必要な分散数を k とすると、秘密にしたい値を切片とするような k-1 次関数を作り、この関数上の n 個の点を分散情報とし、この分散情報を n 台のコンピュータに分散させることにより秘密分散を実現する。ここで、k-1 次関数はこの関数上の k 個以上の点が定まらない限りは一意に定めることができないため、秘密にしたい値は k 個以上の点を集めない限りは復元することができない。この分散方法は、いかなる計算能力を持つコンピュータを使っても秘密にしたい値を復元することはできない、情報理論的安全性という高い安全性を持つ。

本実証研究で用いる秘密計算技術は、複製秘密分散と呼ばれる情報理論的安全性を持つ秘密分散に基づいており、秘密にしたい情報は 3 つに分散され、そのうち 2 つ以上を集めない限りは復元できないことが保証される。

- データの保存：データを複数の分散情報に分け n 台のコンピュータに分散させ保管。
- データの復元：k 台のコンピュータから分散情報を集めてデータを復元。
- 機密性：k-1 台までのコンピュータから分散情報を盗んでも、元のデータに関して何の情報も得られない。
- 可用性：n-k 台まで故障しても、残りのコンピュータからデータを復元できる。

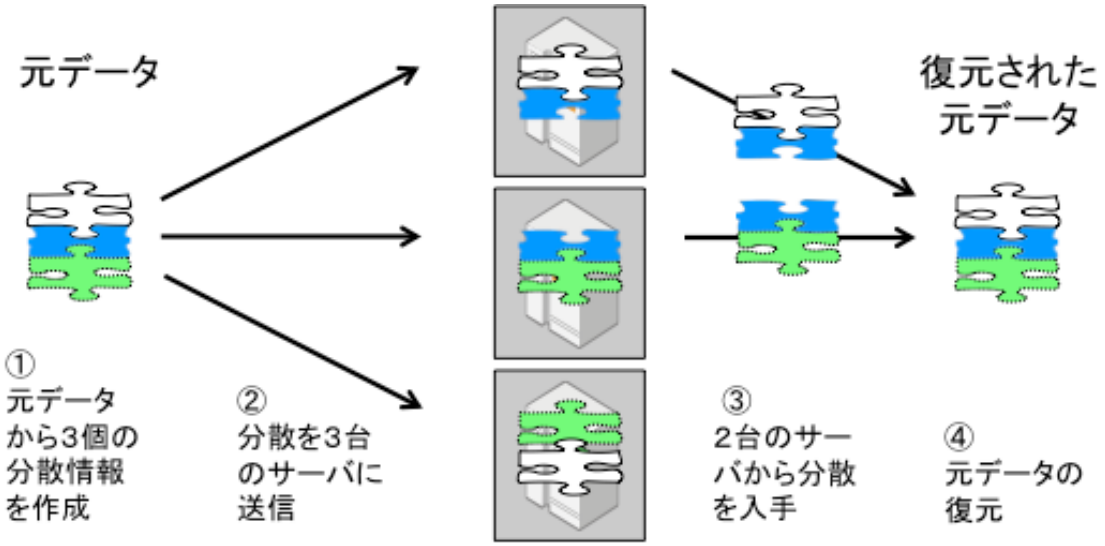


図 3.1 秘密分散の概念図

3.2 秘密計算

3.2.1 秘密計算の概要

秘密計算は、秘密分散を情報の安全性のベースとし、データを秘匿したまま各種の計算を行うことができるシステムである。図は、データを秘匿して計算処理を委託する委託型秘密計算システムの概念図である。入力データは秘密分散されて複数のコンピュータに配置される。秘密計算を構成するコンピュータは、分散データの中身を閲覧することなしに、あらかじめ定められた手順に従ってデータ処理を行い、最終的に求める計算結果を、結果の値の分散として計算出力する。結果は秘密分散と同様に、必要な数の分散情報を集めることによって行われる。

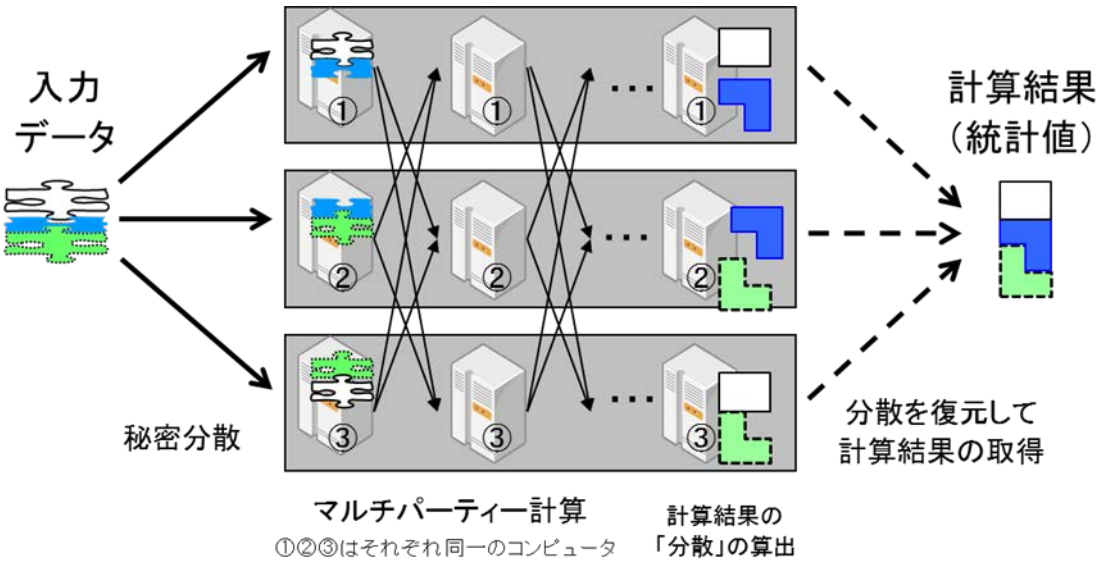


図 3.2 秘密計算の概念図

- データの保存：元データを複数の分散情報に分け n 台のコンピュータに分散させ保管。
- マルチパーティー計算：行うデータ処理に応じて、コンピュータ間で必要なデータ処理とデータの送受信（通信）を必要回数行う。この際、各コンピュータにおいて、他から受信したデータと各コンピュータが保有する分散情報を合わせても各コンピュータは何も元データに関する情報を得られない。
- 機密性：計算過程において各コンピュータは何の情報も得られず、k-1 台までのコンピュータからのデータが集められても何の情報も得られない。
- 結果の出力：計算結果の分散情報が各コンピュータに分散され終了。

秘密計算システムは、上記のマルチパーティー計算環境上に論理演算や算術演算を構築することにより、各種、統計を含む計算が可能になる。

3.2.2 秘密計算の加算

秘密計算システムは、独特の方法で算術演算を実現する。まず始めに秘密計算による加算の実現方法を説明する。

次の図は秘密計算で２つのテストの点（整数）の足し算を行う概念を説明する例である。ここではテストの点は適当な数字（乱数）を使って、3 つの数の和に分散され、秘密分散される。各コンピュータでは、各々が分散データを使ってローカルな加算器として振る舞い、全体の足し算の秘密分散を得る。なお、本実証研究で用いる秘密計算が使う秘密分散はこのような整数の和で表されるものではないが、基本的な概念は同じである。ここで分散値の和が最終的に全体の和になる性質は、秘密分散の加法準同型性という性質が用いられている。

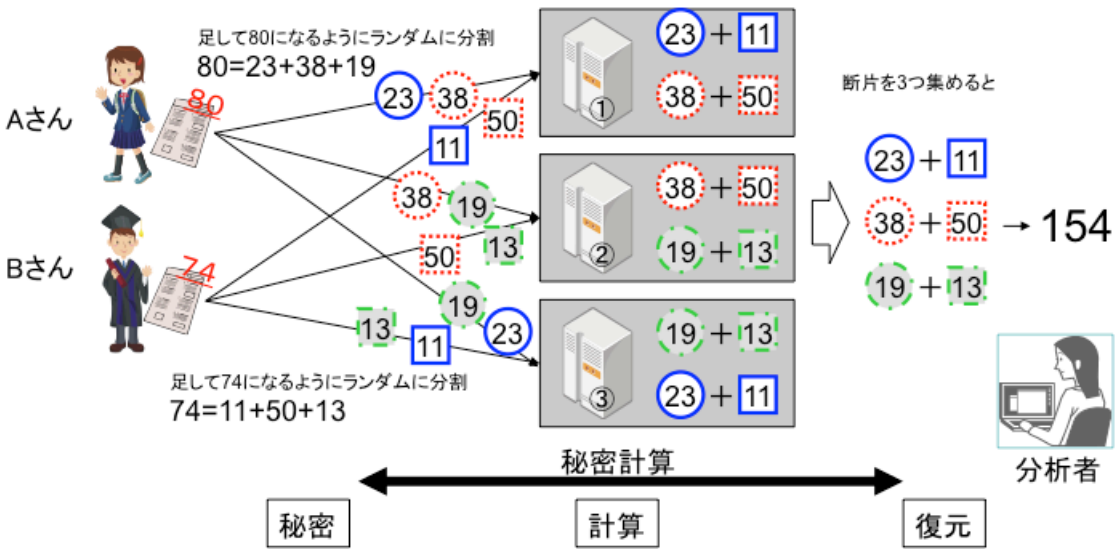


図 3.3 秘密計算の加算

3.2.3 秘密計算の乗算

続いて秘密計算システムにおける乗算の実現方法を説明する。乗算は加算のテクニックを拡張して行う。2 つの整数の積は、図頭に示した、3 つの整数同士の積の和として考えることができる。この積の和の分散値をマルチパーティー計算で求めることが目的となる。図に示した通り、求める部分的な積の和は各コンピュータ内で計算することができる。この積の和を積の秘密分散の形で各コンピュータに配置させるためには、乱数を使ったデータ秘匿を行った通信を行う。以上で乗算が実現できる。

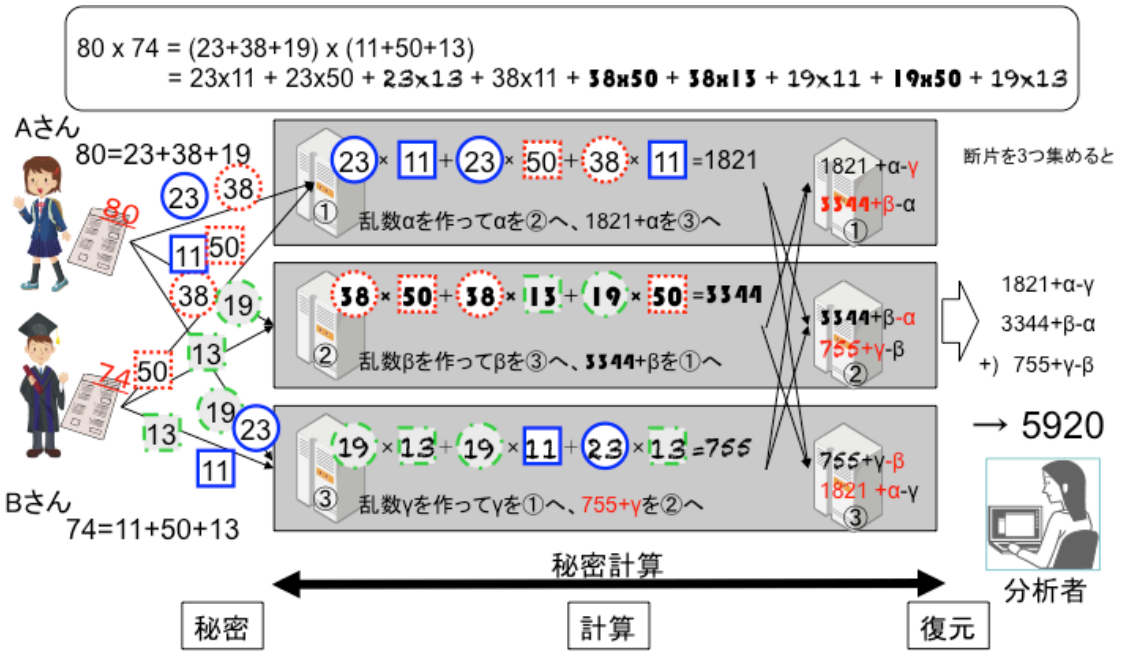


図 3.4 秘密計算の乗算

3.2.4 秘密計算の論理演算

秘密計算システムで実現されている論理演算は、論理積、論理和、排他的論理和、否定、がある。これらは加減算と乗算を用いて実現することができる。

これらの論理演算を組み合わせることで、原理的に任意の計算処理が可能になる。

3.2.5 秘密計算が提供する演算

評価対象とした秘密計算システムでは、以下の演算が提供されている。

- 平均値、分散値、最大値、最小値、中央値
- 条件によるフィルタリング
- カテゴリ別統計値(集計、平均値、分散値、最小値、最大値、中央値)

これらの演算は上述の通り、算術演算や論理演算の組み合わせにより作ることも可能である。しかし、秘密計算システムでは算術演算や論理演算に加えて、ソートや比較などの個別に設計したより高度な演算を組み合わせてこれらの演算を作っている。

その理由は、算術演算や論理演算に基づいた一般的な構成方法で作ると多くの演算で処理性能が低くなってしまうからである。秘密計算では、動作の違いによって本来明かしてはならないデータについての情報が漏れることを防ぐため、入力や計算途中のデータに応じて動作を変えることは許されない。このため、一般的な構成方法では、情報を漏らさないために、条件分岐があった場合は両方の場合を計算して本来の分岐先の結果だけを残すという処理が行われる。その結果、条件分岐が多く含まれる演算を一般的な構成方法を使って秘密計算で実現すると計算時間が非常に大きくなってしまう。そこで、秘密計算ではいくつかの重要な演算を個別に効率よく実現し、これらを部品として用いて演算を実装している。

3.2.6 秘密計算のソート

秘密計算システムで使われている個別に設計された演算の一つにソートがある。ソートは入力として受け取った値の列を昇順に並び替える演算であり、通常の計算機上でも多くの処理の部品として使われる重要な演算である。秘密計算システムにおいても、最大値や最小値、中央値、カテゴリ別統計値の計算において、ソートは中心的な役割を果たす。秘密計算におけるソートは、秘密分散された値の列を入力とし、別の秘密分散された値の列を出力する処理であり、出力を復元した列が入力を復元して昇順に並べ変えた列と一致するように計算される。

秘密計算システムでは、公開しても入力に関する情報が一切漏れない中間データを作りだし、この中間データに依存して動作を変える工夫により、一切動作を変えないために非常に計算時間が大きくなってしまう一般的な構成方法に比べて効率よくソートを実現している。具体的には、以下の手順で処理を行う。

1. 各要素のソート後の順位を計算
2. 全体をランダム置換
3. 順位だけを復元
4. 復元した順位に従って移動

ソート後の順位の計算は、算術演算と論理演算の組み合わせにより効率よく行われる。ランダム置換は、秘密計算用に個別に設計された演算であり、秘密分散された値の列を入力として、出力の秘密分散された値を復元した列が入力の秘密分散された値を復元してランダムな順序に並び替えた列となっているように、別の暗号文の列を計算して出力する処理である。ランダム置換を用いて誰にもわからない順序に並び替えることにより、計算された順位を公開しても入力に関する情報は一切漏れない。最後に順位を復元し、これを利用して昇順への並べ替えを実現する。このように一般的な構成方法では必要となる条件分岐を回避したことで、ソートが効率よく実現される。

4 秘密分散･秘密計算技術の集計表作成ソフトウェアへの組み込みの実証・性能評価

情報安全性を確保した統計データ提供の一形態と考えられるリモートアクセス型オーダーメード集計の実用化に資する技術的な実証のため、集計表作成ソフトウェアに秘密分散・秘密計算技術を組み込み、組み込み後ソフトウェアのデータ処理に関する性能評価を行った。

4.1 集計表作成ソフトウェアへの秘密分散･秘密計算技術の組み込み

秘密分散・秘密計算技術を汎用集計ソフトウェアに組み込むにあたっては、以下の観点から方式の検討を行った。

- 既存の資産を可能な限り活用可能であること
- 今後の変更や拡張に柔軟に対応できるよう、疎結合で開発を行えること

検討の結果、秘密分散・秘密計算技術と汎用集計ソフトウェアである Adam-Lite を連係させる組み込み方式として、新たに開発する専用パイプモジュールを用いてデータの受け渡しを行う方式を採用した。

4.1.1 組み込みイメージ

専用パイプモジュールを使用した秘密分散・秘密計算技術の汎用集計ソフトウェアへの組み込みイメージを以下に示す。専用パイプモジュールを使用した組み込み方式では、集計クライアントから集計が指示されると、秘密計算サーバが必要な処理を行い、接続用のパイプモジュールはその結果のデータを秘密分散・計算モジュールから受け取り、Adam-Lite に対してデータを受け渡し、Adam-Lite により集計表が作成される。

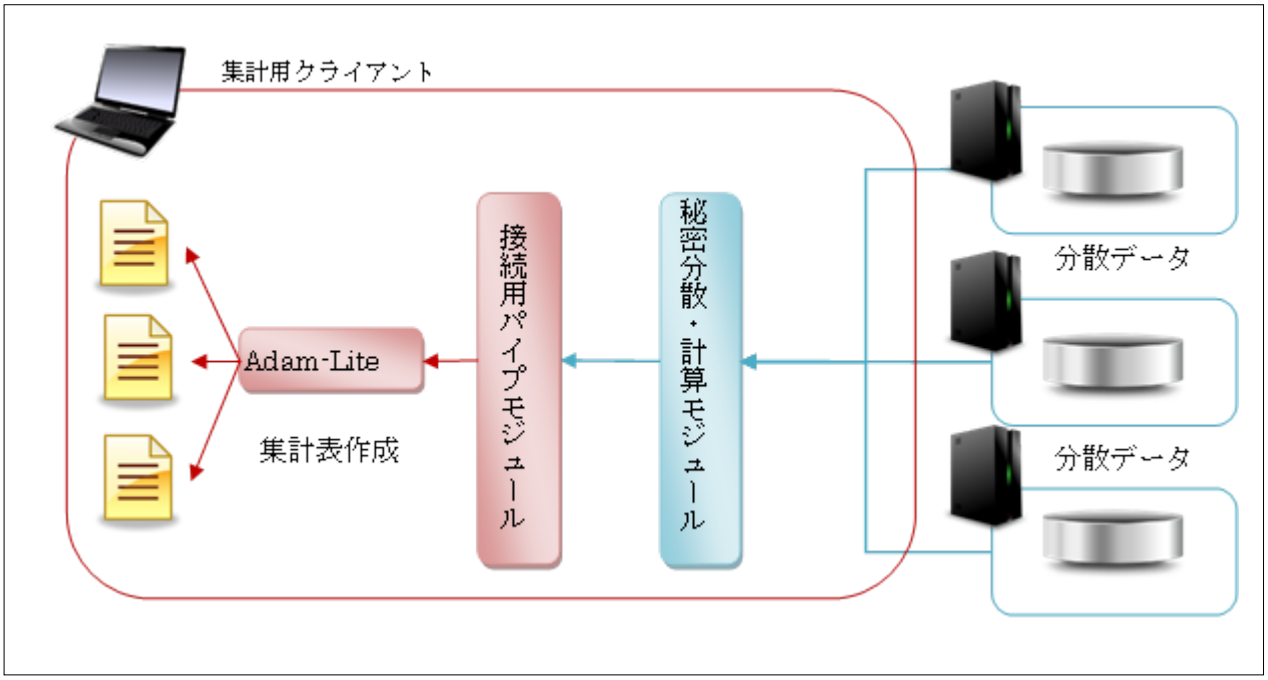


図 4.1　接続用パイプモジュールを使用した組み込みイメージ

4.1.2 処理フロー

秘密分散・秘密計算モジュールは、Adam-Lite からの集計指示に従い、下記のフローに沿って復元、及び秘密計算を行う。

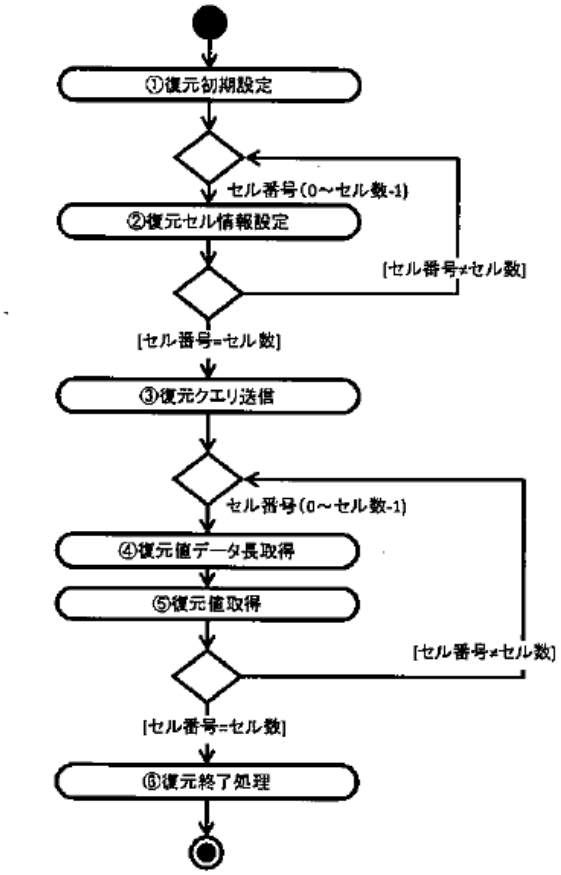


図 4.2 秘密分散データの復元フロー

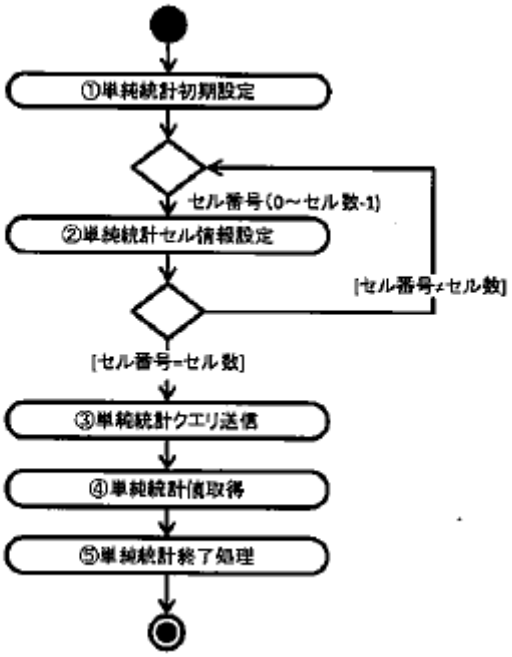


図 4.3 秘密計算処理フロー

4.1.3 Adam-Lite からの利用

Adam-Lite から秘密分散・秘密計算モジュールを使用する際は以下のファイルを準備し、Adam-Lite スクリプトに接続用パイプモジュール使用を宣言する。

- 秘密計算モジュール設定ファイル、認証用鍵
- Adam-Lite にて接続用パイプモジュール設定ファイル

4.1.3.1 秘密計算モジュール設定ファイル

```
## 秘密計算評価プログラム クライアント設定ファイル
## 秘密計算サーバ数（3 以上）
SERVER_NUM = 3

## 秘密計算サーバ 1
SERVER_NAME = 192.168.1.155
SERVER_PORT = 6001
SERVER_CERT = cert_server.der

## 秘密計算サーバ 2
SERVER_NAME = 192.168.1.155
SERVER_PORT = 6002
SERVER_CERT = cert_server.der

## 秘密計算サーバ 3
SERVER_NAME = 192.168.1.155
SERVER_PORT = 6003
SERVER_CERT = cert_server.der
```

図 4.4 秘密分散・計算モジュール設定ファイル

4.1.3.2　Adam-Lite 指定方法

Adam-Lite では外部連係用のパイプを DATA コマンド内に記述する事ができる。下記の例では Pipe_scsvr という名の接続用パイプモジュールを DATA コマンド内で指定している。

```
DATA PIPE#PIPE_SCSVR tablename [-P
pipe_name] [-I info_file]
```

図 4.5 パイプモジュール指定方法

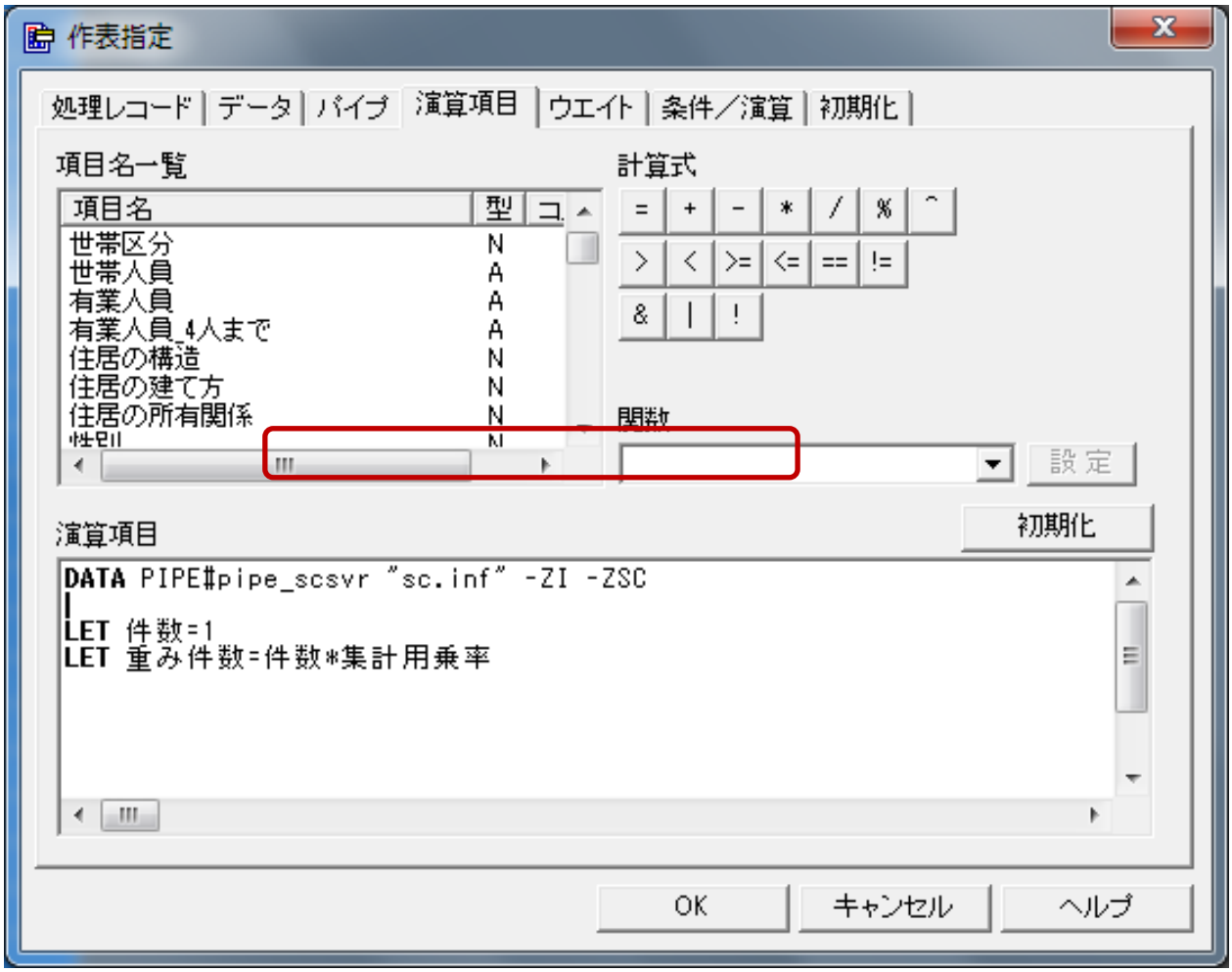


図 4.6 接続用パイプモジュール指定画面

4.2 評価

本節では、専用パイプモジュールを使用した組み込み方式について、実際に秘密分散・秘密計算技術を使用したデータを集計し、目的とする集計表の作成が可能であることの実証及びその処理性能についての評価を実施する。

4.2.1 対象データ、評価項目

専用パイプモジュールを使用した組み込み方式の検証に使用するデータと評価対象を以下に示す。

- データ：擬似ミクロデータ（平成 16 年全国消費実態調査）
- 評価対象：疑似ミクロデータのクロス集計表

4.2.1.1 データ概要

擬似ミクロデータで提供される 32,027 レコードを対象に評価を行う。

表 4.1 疑似ミクロデータ概要

レコード数	32,027 レコード 二人以上の勤労者世帯	
ファイルサイズ	giji_2004zensho_dataset(00001～05000).csv giji_2004zensho_dataset(05001～10000).csv giji_2004zensho_dataset(10001～15000).csv giji_2004zensho_dataset(15001～20000).csv giji_2004zensho_dataset(20001～25000).csv giji_2004zensho_dataset(25001～30000).csv	約 15MB
	giji_2004zensho_dataset(30001～32027).csv	約 6MB
世帯属性等	14 項目　世帯主の年齢、住居の種類など	
支出項目	149 項目 用途分類	
収入項目	34 項目　年間収入等	

4.2.1.2 データレイアウト

下記に擬似ミクロデータのレイアウト（抜粋）を示す。

表 4.2　疑似ミクロデータレイアウト（抜粋）

政府統計コード	00200564	実施時期	平成１６年	作成日	平成23年8月25日	訂正日	平成24年8月9日
統計調査名	全国消費実態調査	集計区分	２人以上勤労世帯	ファイル名			
調査票名	擬似ミクロデータ（大規模データ）	（備考・補足事項）			（照会先等）		
コード体系	Shift_JIS	「△」はブランク（空白）を示す。					
レコード長	2974						

行番号	項目名	階層	位置	バイト数	配置	型	種別	変数名	符号	符号内容	備考
1	世帯区分	1	1	1				SetaiKubun	1	勤労	
2									2	勤労以外	
3									3	無職	
4	,	1	2	1					,	項目の区切り	
5	世帯票	1									
6	世帯人員	2	3	2	2	1		SetaiJinin	△2～	２人～	
7	,	2	5	1					,	項目の区切り	
8	有業人員	2	6	2	2	1		ShuugyouJinin	△1～	１人～	
9									VV	不詳	
10	,	2	8	1					,	項目の区切り	
11	現住居等に関する事項	2									
12	住居の構造	3	9	1				Kouzou	1	木造	
13									2	防火木造	
14									3	鉄骨・鉄筋コンクリート造	
15									4	その他（ブロック造り、レンガ造りなど）	
16									V	不詳	
17	,	3	10	1					,	項目の区切り	
18	住居の建て方	3	11	1				Tatekata	1	一戸建	
19									2	長屋建	
20									3	共同住宅（１・２階建）	
21									4	共同住宅（３～５階建）	
22									5	共同住宅（６～１０階建）	
23									6	共同住宅（１１階建以上）	
24									7	その他	
25									V	不詳	
26	,	3	12	1					,	項目の区切り	

4.2.1.3　擬似ミクロデータサンプル

以下に擬似ミクロデータのサンプル（抜粋）を示す。

```
1, 2, 1,1,1,1,1, 3,1,1,2,VV, 3,        16.90000,      5853.01640,    649738.97217,    339553.25922,    295731.13659,  29
1, 2, 1,1,1,1,1, 3,1,1,2,VV, 3,         9.93333,      2207.21640,    568988.01380,    280663.05945,    244441.19830,  24
1, 2, 1,1,1,1,1, 3,1,1,2,VV, 3,        10.96667,      2248.30560,    454047.55185,    221530.64635,    192940.30629,  19
1, 2, 1,1,1,1,1, 3,1,1,2,VV, 3,        14.10000,      3947.41130,    453147.62861,    285956.11755,    249051.14400,  24
1, 2, 1,1,1,1,1, 3,1,1,3, 9, 1,        17.20000,      4071.80030,    584970.67321,    258830.20596,    255514.30560,  18
1, 2, 1,1,1,1,1, 3,1,1,3, 9, 1,        13.60000,      4391.75330,    917737.39188,    646259.12451,    637979.83246,  45
1, 2, 1,1,1,1,1, 3,1,1,3, 9, 1,        18.16667,      4746.53790,    974959.97538,    214521.06124,    211772.80988,  15
1, 2, 1,1,1,1,1, 3,1,1,3,VV, 3,        15.31111,      4077.62390,    325183.81328,    285336.97129,    269278.39879,  26
1, 2, 1,1,1,1,1, 3,1,1,3,VV, 3,        13.86667,      4408.44040,    253708.30377,    131342.75858,    123950.87662,  12
1, 2, 1,1,1,1,1, 3,1,1,3,VV, 3,        15.31111,      7343.94400,   1588918.45354,    339983.61366,    320849.56494,  32
1, 2, 1,1,1,1,1, 3,1,1,3,VV, 3,        14.40000,      3236.44360,    270761.53550,    156006.11949,    147226.20019,  14
1, 2, 1,1,1,1,1, 3,1,1,3,VV, 3,        15.31111,      5624.24840,    838560.42277,    452658.38132,    427183.07259,  42
1, 2, 1,1,1,1,1, 3,1,1,5,VV, 3,        17.06667,      5340.29970,    554976.77651,    335253.91035,    332853.81796,  24
1, 2, 1,1,1,1,1, 3,1,1,5,VV, 3,        17.06667,      6000.22140,    672072.60335,    434492.36251,    431381.81919,  31
1, 2, 1,1,1,1,1, 3,1,1,5,VV, 3,         6.26667,      5844.75090,    463014.45313,    315049.92219,    312794.47073,  22
```

図 4.7 擬似ミクロデータサンプル

4.2.1.4　評価対象

専用パイプモジュールを使用した組み込み方式の実証における評価対象として、疑似ミクロデータのクロス集計表を使用する。出力サンプルとして提供される以下の表 4.3、表 4.4、表 4.5、表 4.6 の 4 表について評価を行う。

表 4.3 は有業人員×世帯人員というクロス表であり、単純にカウント集計した表である。表 4.4 は各レコードを集計用乗率で重み付けをしてカウントした表である。これらの表について、表 4.7 に示した 3 パターンについて性能測定を行った。

表 4.5 は表 4.4 を元に、総数を 10 万に調整した表である。表 4.5 については作成されたクロス表、表 4.3、表 4.4 を元に作成が可能であり、元データからのアクセスが必要ない事から、パフォーマンスの計測から除外した。

表 4.6 については、表の各セルごとの統計値の計算を秘密計算で計算した場合の処理時間を測定し、4.2 節で述べるパターン 3 で表全体を処理した場合の処理時間を見積もった。

表 4.3 集計世帯数(各レコードを、単純にカウントしたもの)

		総数	世帯人員								
			2人	3人	4人	5人	6人	7人	8人	9人	10人
総数		32,027	7,438	8,537	9,944	4,405	1,214	390	81	15	3
有業人員	1人	13,913	4,124	3,908	4,132	1,436	256	51	6	0	0
	2人	13,459	3,239	3,391	4,201	1,943	494	162	29	0	0
	3人	2,950	0	1,035	1,031	559	232	84	6	3	0
	4人	691	0	0	324	220	104	31	12	0	0
	5人	40	0	0	0	27	6	7	0	0	0
	6人	6	0	0	0	0	3	3	0	0	0
	不詳	968	75	203	256	220	119	52	28	12	3

表 4.4 世帯数分布(各レコードを、集計用乗率で重み付けして、カウントしたもの)

		総数	世帯人員								
			2人	3人	4人	5人	6人	7人	8人	9人	10人
総数		495,465	115,835	132,005	155,850	67,565	17,161	5,611	1,197	207	33
有業人員	1人	219,743	64,691	61,284	66,299	22,740	3,813	831	84	0	0
	2人	205,723	50,138	51,523	64,868	29,616	6,801	2,336	441	0	0
	3人	44,041	0	15,912	15,615	7,963	3,302	1,137	76	36	0
	4人	10,168	0	0	4,851	3,345	1,354	422	195	0	0
	5人	570	0	0	0	383	80	108	0	0	0
	6人	99	0	0	0	0	49	51	0	0	0
	不詳	15,120	1,006	3,287	4,216	3,519	1,761	727	401	170	33

表 4.5 世帯数分布(10 万比)

		総数	世帯人員								
			2人	3人	4人	5人	6人	7人	8人	9人	10人
総数		100,000	23,379	26,643	31,455	13,637	3,464	1,132	242	42	7
有業人員	1人	44,351	13,057	12,369	13,381	4,590	770	168	17	0	0
	2人	41,521	10,119	10,399	13,092	5,977	1,373	471	89	0	0
	3人	8,889	0	3,211	3,152	1,607	666	229	15	7	0
	4人	2,052	0	0	979	675	273	85	39	0	0
	5人	115	0	0	0	77	16	22	0	0	0
	6人	20	0	0	0	0	10	10	0	0	0
	不詳	3,052	203	664	851	710	355	147	81	34	7

表 4.6 支出(消費支出 及び 十大費目、単位:円)

			集計世帯数	世帯数分布(抽出率調整済)	消費支出	食料	住居	光熱・水道	家具・家事用品	被服及び履物	保健医療	交通・通信	教育	教養娯楽	その他の消費支出
総数			32,027	495,465	328,140	72,883	17,687	19,238	9,204	14,138	11,366	47,961	22,270	31,389	82,003
	うち世帯人員が４人		9,944	155,850	335,438	76,362	15,345	20,214	8,885	14,452	10,987	47,894	33,442	32,269	75,588
	有業人員	１人	4,132	66,299	305,234	71,543	17,556	18,854	8,383	13,579	11,656	42,703	31,202	31,959	57,801
		２人	4,201	64,868	347,740	78,472	12,932	20,621	8,917	14,876	10,538	52,141	39,485	33,681	76,078
		３人	1,031	15,615	380,521	83,796	12,583	23,045	10,276	16,561	10,331	52,406	22,513	27,852	121,160
		４人	324	4,851	399,962	85,083	18,500	22,490	11,763	14,096	10,812	51,310	4,509	32,769	148,628
		不詳	256	4,216	379,882	82,134	24,296	22,238	7,843	14,238	10,011	43,521	49,453	31,206	94,942
		(特掲) １人又は２人	8,333	131,168	326,256	74,970	15,269	19,728	8,647	14,221	11,103	47,371	35,298	32,810	66,839

4.2.2 評価方法

接続用パイプモジュールを使用した組み込み方式の実証における評価ケースとして、3 つのパターンを用いて評価を行う。その結果、パターン 1、パターン 2、パターン 3 で作成される集計表が同じ結果である事を確認する。また、それぞれのパターンでのパフォーマンスを計測し、比較する。

パフォーマンス計測では計測環境の状態による影響を排除する為に 3 度の計測を行い、平均値を採用する。

- パターン 1 : 秘密計算モジュールを使用しないで集計（従来通りの運用）
- パターン 2 : 秘密分散された状態のデータから明細データ取得（従来より安全性の高い方式）
- パターン 3 : 秘密計算後の結果のみを取得（パターン 2 よりさらに安全性の高い方式）

また、各パターンでは以下のような役割分担で処理を行う。

表 4.7 各パターンでの役割分担

	パターン 1	パターン 2	パターン 3
データ取得	Adam-Lite	秘密分散	秘密分散
データ集計	Adam-Lite	Adam-Lite	秘密計算
表演算	Adam-Lite	Adam-Lite	Adam-Lite
製表処理	Adam-Lite	Adam-Lite	Adam-Lite
レイアウト処理	Adam-Lite	Adam-Lite	Adam-Lite

データ集計後の表演算、製表処理、レイアウト処理の各処理は、一般に生データに対して直接処理を行うのではなく、集計後のオブジェクトとして汎用集計ソフトウェアで行う場合が多い。

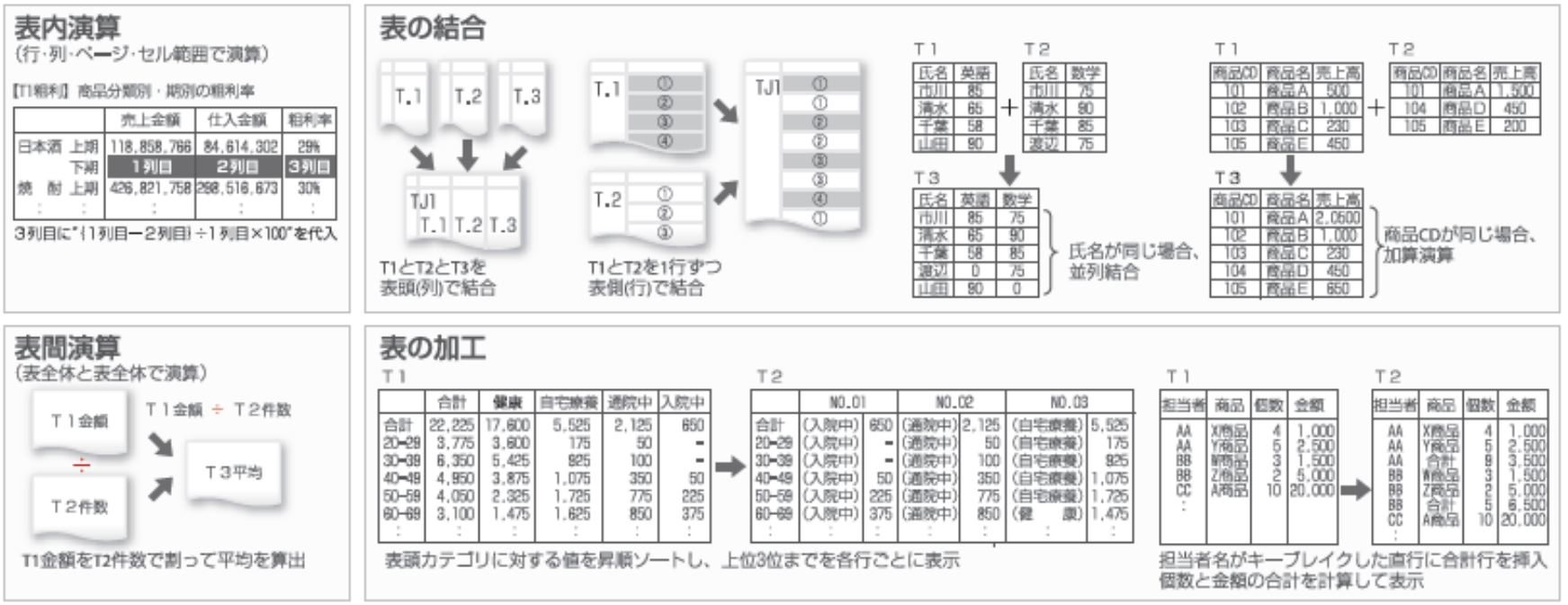


図 4.8 Adam-Lite の表演算機能

今回の評価対象では、表 4.5 が表 4.4 から計算可能であり、元データから集計を行うのではなく表間演算により出力するのが適切な例となっている。特に Adam-Lite は表内、表間演算や表の結合や表の加工を行う為の機能を豊富に揃えており、集計後の処理は Adam-Lite に割り当てるのが合理的である。したがって、表演算以降のプロセスは Adam-Lite が行い、それ以前の集計処理までを 3 パターンで実行、比較を行うこととした。

4.2.2.1　パターン 1：秘密計算モジュールを使用しないで集計

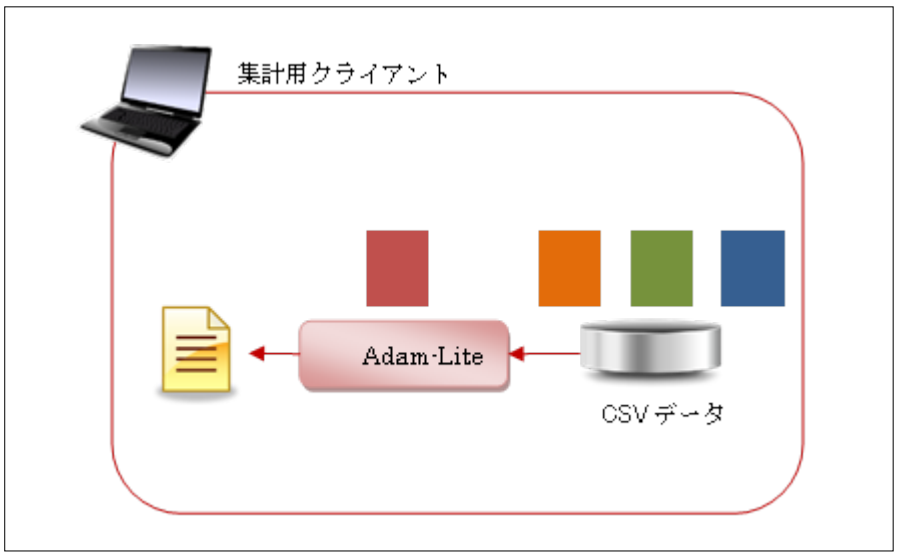


図 4.9 パターン 1 概要図

パターン 1 は秘密分散・秘密計算技術は使っていない、従来型の運用と一致する形態である。単一の PC 内ですべての処理が行われる。集計に用いるデータは単一のデータベース内で保管されており、後述のパターンに比べると安全性は低い。

4.2.2.2　パターン 2：　秘密分散された状態のデータから明細データを取得

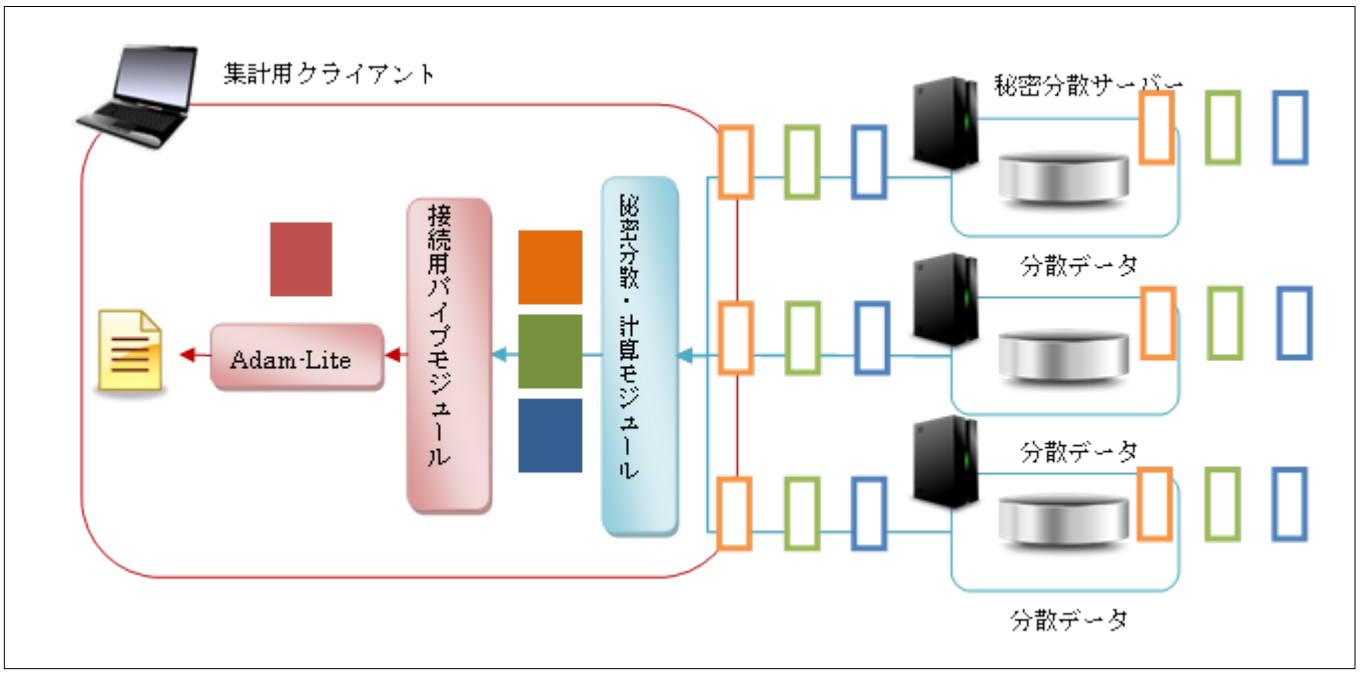


図 4.10　パターン 2 概要図

パターン 2 は秘密分散技術を利用した形態である。データは秘密分散技術を用いて複数の秘密分散サーバに分割保管される。集計を行う際は、集計に必要なデータに対応するデータのみが集計用クライアントに復元され、所望の集計処理が行われる。集計の対象のみが、集計の際にだけ復元されるため、パターン 1 に比べて安全性は高い。

4.2.2.3　パターン 3: 秘密計算後の結果のみを取得

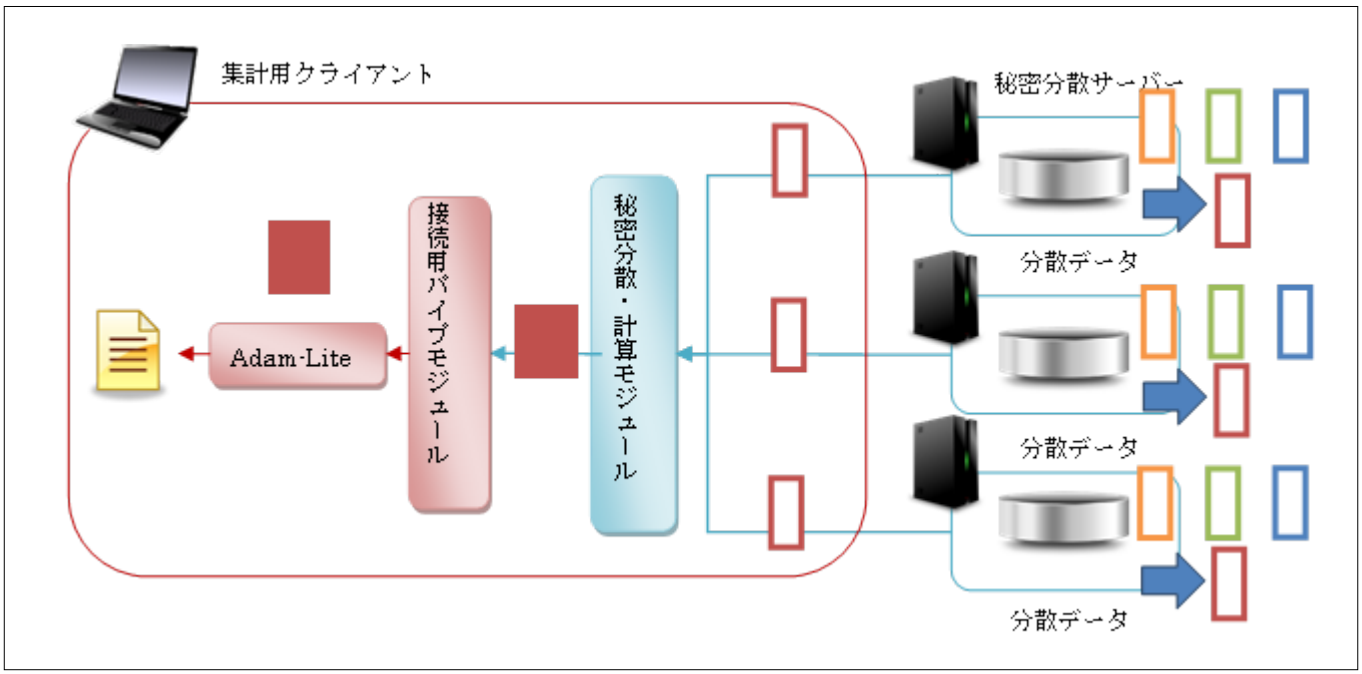


図 4.11　パターン 3 概要図

パターン 3 は秘密分散技術に加え、秘密計算技術も利用した形態である。パターン 2 と同様にデータは秘密分散サーバに分割保管され、さらに集計の処理も秘密分散サーバ間で秘密計算技術により元のデータや途中の計算結果を一度も復元することなく行われる。最終的な出力となる集計結果のみが集計用クライアントに復元されるため、極めて安全性が高い。

4.2.2.4　評価環境

評価環境は以下に記す通りである。本評価では集計用クライアントと 3 台の秘密分散サーバを 1Gbps の LAN により接続し、測定を行った。

表 4.8　集計用クライアント

CPU	Intel Core i7-2600 3.40GHz
RAM	16GB
OS	Windows 7 Professional
SSD	128GB

表 4.9　秘密分散サーバ(3 台)

CPU	Intel Core i7-2640M 2.80GHz
RAM	8GB
OS	Linux(Ubuntu 12.04)
SSD	128GB

4.2.3　評価結果

検証の結果、接続用パイプモジュールを使用する事で、秘密分散・秘密計算モジュールから Adam-Lite への接続及び評価対象の帳票を出力する事ができた。出力された帳票の数値は小数点以下を四捨五入した整数値において全て同じ物である事が確認でき、処理が正しく行われている事も確認できた。次に処理パフォーマンスに関して以下に記す。

4.2.3.1　表 4.3 の処理時間

表 4.10　表 4.3 の処理時間比較表（秒）

	平均	1 回目	2 回目	3 回目
パターン 1	2.82	2.84	2.81	2.82
パターン 2	3.28	3.18	3.48	3.17
パターン 3	182.22	182.39	182.48	181.80

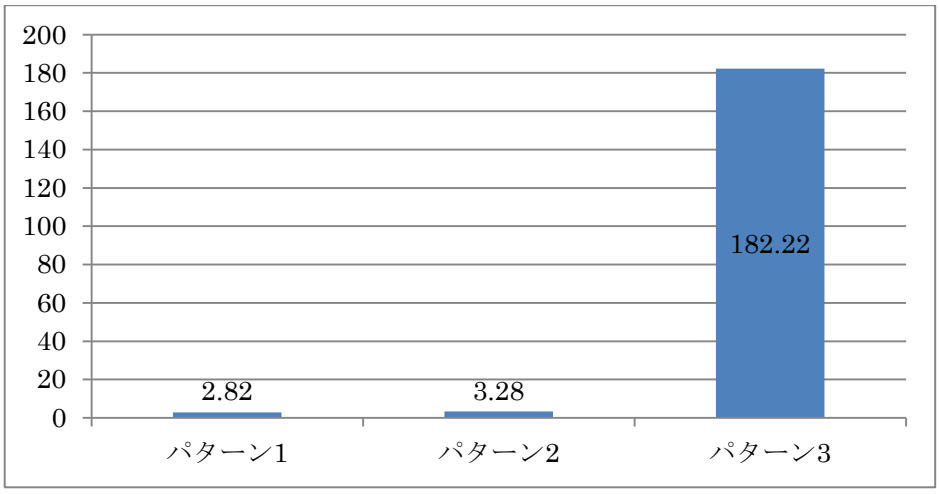


図 4.12　表 4.3 の処理時間比較（秒）

4.2.3.2　表 4.4 の処理時間

表 4.11 表 4.4 の処理時間比較表（秒）

	平均	1 回目	2 回目	3 回目
パターン 1	2.83	2.82	2.86	2.82
パターン 2	3.26	3.18	3.12	3.46
パターン 3	364.32	365.42	364.27	363.28

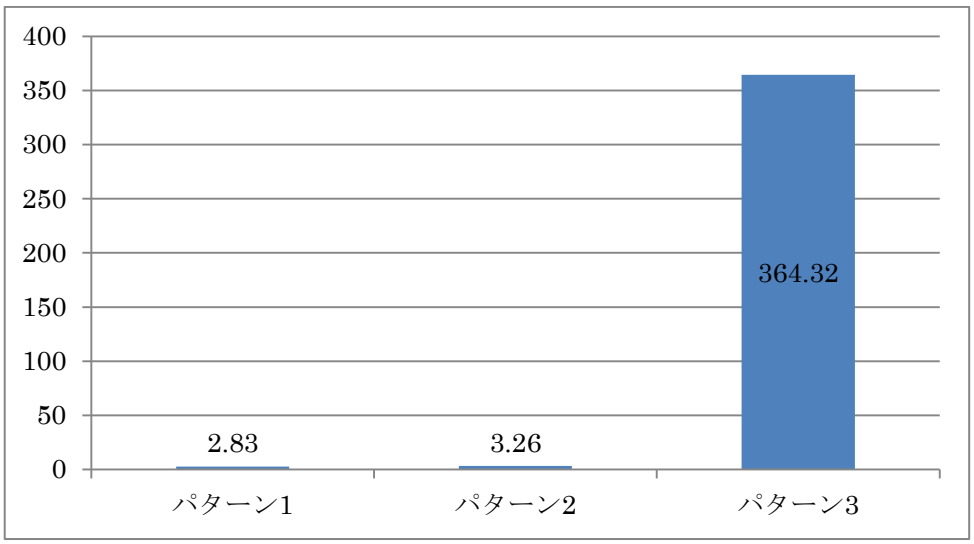


図 4.13　表 4.4 の処理時間比較（秒）

処理時間に関しては、秘密分散・秘密計算モジュールを使用しないパターン 1 が最も高速に処理する事ができ、次にパターン 2、パターン 3 の順番となった。パターン 2 は秘密分散のみを行っている事から、パターン 3 との比較により秘密計算の処理に多くの時間が掛かっている事がわかる。データの機密性はパターン 2 に比べパターン 3 の方が高く、機密性とパフォーマンスはトレードオフの関係にある事が確認できる。特に表 4.4 は単純集計に加えて、重み付けを行う集計結果であった為、更に違いが顕著に現れた。

4.2.3.3　表 4.6 のパターン 3 による処理時間の見積もり

表 4.6 の各セルの統計値をそれぞれ秘密計算で計算した場合の処理時間を測定し、表 4.6 全体をパターン 3 で処理した場合の処理時間を見積もった。

表 4.6 のうち、灰色に塗りつぶしたセルの統計値は、表の中の別の統計値または別の表の統計値から計算できるため、測定対象から除外した。集計世帯数の列の値(表 4.12 中のa_1からa_8)は表 4.3 から、世帯数分布の列(表 4.12 中のb_1からb_8)は表 4.4 から、世帯人員が 4 人の平均値は世帯数分布による加重平均(例えば$c_2 = (\sum_{i=3}^{7} b_i c_i)/b_2$)により、1 人又は 2 人平均値は世帯数分布による加重平均(例えば、$c_8 = (b_3 c_3 + b_4 c_4)/(b_3 + b_4)$)により、それぞれ計算できる。

表 4.6 の統計値を秘密計算により計算した場合の処理時間を表 4.13 および表 4.14 に示す。表 4.13 は、消費支出からその他の消費支出までの各列の総数の計算をした場合の処理時間である。表 4.14 は、消費支出からその他の消費支出までの列ごとに有業人員別(1 人、2 人、3 人、4 人、不詳)の平均値を一括で計算した場合の処理時間である。表全体で秘密計算に要した処理時間は、3 回の平均で 1060.57 秒であり、表 4.6 をパターン 3 によって計算する場合の秘密計算に要する処理時間はおおよそ 1060 秒程度と見積もられる。

表 4.12 表中の値から計算可能な統計値の例

			集計世帯数	世帯数分布(抽出率調整済)	消費支出	食料
総数			a_1	b_1	c_1	d_1
	うち世帯人員が４人		a_2	b_2	$c_2 = (\sum_{i=3}^{7} b_i c_i)/b_2$	$d_2 = (\sum_{i=3}^{7} b_i d_i)/b_2$
	有業人員	１人	a_3	b_3	c_3	d_3
		２人	a_4	b_4	c_4	d_4
		３人	a_5	b_5	c_5	d_5
		４人	a_6	b_6	c_6	d_6
		不詳	a_7	b_7	c_7	d_7
		(特掲) １人又は２人	$a_8 = a_3 + a_4$	$b_8 = b_3 + b_4$	$c_8 = \frac{b_3 c_3 + b_4 c_4}{b_3 + b_4}$	$d_8 = \frac{b_3 d_3 + b_4 d_4}{b_3 + b_4}$

表 4.13 総数の秘密計算による計算の処理時間(秒)

	消費支出	食料	住居	光熱・水道	家具・家事用品	被服及び履物	保健医療	交通・通信	教育	教養娯楽	その他の消費支出
1回目	0.52	0.89	0.56	0.48	0.69	0.68	0.63	0.50	0.70	0.50	0.51
2回目	0.48	0.48	0.64	0.48	0.49	0.51	0.52	0.49	0.65	0.75	0.79
3回目	0.49	0.49	0.50	0.49	0.93	0.81	0.50	0.51	0.48	0.50	0.51
平均	0.50	0.62	0.57	0.49	0.71	0.66	0.55	0.50	0.61	0.58	0.60

表 4.14 有業人員別の平均値の秘密計算による計算の処理時間(秒)

	消費支出	食料	住居	光熱・水道	家具・家事用品	被服及び履物	保健医療	交通・通信	教育	教養娯楽	その他の消費支出
1回目	95.43	96.01	95.81	95.29	95.92	96.09	95.73	95.80	95.63	96.03	95.92
2回目	96.06	95.80	95.72	96.22	95.97	96.23	95.90	96.34	96.42	96.40	96.65
3回目	95.09	95.34	95.52	95.62	95.38	95.76	95.60	96.29	95.39	95.85	95.37
平均	95.53	95.72	95.68	95.71	95.76	96.02	95.74	96.15	95.81	96.09	95.98

5 リモートアクセス型オーダーメード集計の実用化に向けた技術的課題

本節では、リモートアクセス型オーダーメード集計の実用化に向けた、今後の技術的課題について議論する。

5.1 処理時のパフォーマンス

本研究では 3 万件程度のデータでパフォーマンス計測を行ったが、通常の集計時間に比べて秘密分散、秘密計算モジュールを経由する事でのオーバーヘッドが確認された。公的データの内容によっては更にデータ件数が増える事が予想され、今後のパフォーマンス改善が求められる。

パフォーマンスの改善方法の一つは、秘密計算上の演算の処理速度の改良である。本研究で用いた多くの演算は高速化の研究の途上にあり、例えば本研究で使った実装に含まれているソートや比較などの基本的な演算についても既により高速な新しいアルゴリズムが提案されている。このように、今後も演算の改良によるパフォーマンスの向上が見込まれる。

また、統計処理に向いた秘密計算上の演算の設計もパフォーマンスの有力な改善方法である。例えば、本研究では集計表の作成の際にカテゴリごとの集計や総和の秘密計算をそれぞれ行っていたが、これらの計算は集計表作成の多くの場合に一緒に行われる。そこで、これら両方の計算結果を一括で計算する専用の効率の良い演算を設計できれば、集計表の作成を行う多くの場合にパフォーマンス向上が期待できる。

5.2 秘密計算と汎用集計ソフトウェアとの処理の分担

秘密分散・秘密計算技術を使用した場合、既に集計された結果の数値として、汎用集計ソフトウェアにデータが引き渡される。しかし、重み付け集計や、複雑な多重クロス表作成、表同士の演算などは汎用集計ソフトウェアで行う方が望ましいケースも存在する。特に計算量が多い処理や、多重クロス表などは処理パフォーマンスの問題も存在する。

必要とされるパフォーマンスや安全性、あるいは開発の効率によって汎用集計ソフトウェアと秘密計算モジュールでの処理内容をどう分担させるかを見極める必要がある。

特に、表の知識の利用は、パフォーマンス改善に有効である。表 5.1 に表の知識が有効な表の例を示す。この表においては、(集計 1)の部分はセルごとに集計値を計算する必要があるが、(総数 1)および(総数 2)の各セルの値は(集計 1)の集計値から、(総数 3)の値は(総数 1)または(総数 2)の値から、それぞれ計算することができる。従って、全体を秘密計算で計算する代わりに(集計 1)の部分のみを秘密計算で計算し、その後汎用集計ソフトウェア上で上述の知識を使って(総数 1)、(総数 2)、(総数 3)の部分をそれぞれ計算することで、パフォーマンスを改善することができる。例えば、4.2.3.1 節のパターン 3 で行った処理は、(集計 1)に相当する部分の秘密計算に 90.10 秒(全体の 49.44%)、(総数 1)および(総数 2)に相当する部分の秘密計算にそれぞれ 45.97 秒(全体の 25.23%)、45.90 秒(全体の 25.19%)の処理時間がかかっている。表の知識の利用により(総数 1)および(総数 2)の秘密計算に要する時間を削減することで、全体で約 2 倍の高速化が見込まれる。

表 5.1 知識の利用が有効な表の例

<table>
<tr><td colspan="2" rowspan="2"></td><td rowspan="2">総数</td><td colspan="5">世帯人員</td></tr>
<tr><td>２人</td><td>３人</td><td>４人</td><td>５人</td><td>６人以上</td></tr>
<tr><td colspan="2">総数</td><td>（総数 3）</td><td colspan="5">（総数 2）</td></tr>
<tr><td rowspan="4">有業人員</td><td>１人</td><td rowspan="4">（総数 1）</td><td colspan="5" rowspan="4">（集計 1）</td></tr>
<tr><td>２人</td></tr>
<tr><td>３人</td></tr>
<tr><td>４人以上</td></tr>
</table>

5.3 秘密計算と汎用ソフトウェアのシームレスな連動

安全性の観点では秘密計算で可能な限りの集計処理を行った上で汎用集計ソフトウェアにデータを引き渡す方法が望ましい。ただし、その為には汎用集計ソフトウェアと秘密計算システムとの間で、集計単位情報や項目情報等を状況に応じて、的確に連係させる必要がある。本研究では実験的に固定的な情報を引き渡して計測を行ったが、将来的にはこれらの情報を指定する為の専用 GUI や、自動連動方法などの検討が必要になってくる。

5.4 秘匿処理

本実験ではデータの保管から集計処理にかけての機密性の確保については検証を行う事ができた。しかしながら、公的データの二次活用においては、出力されるデータそのものの秘匿処理が必要となる。秘匿処理に関しては、技術的なアプローチの他に、運用面からのアプローチも検討が必要である。

5.5 認証

リモートアクセス型のオーダーメード集計の実現に向けて、利用者の認証も課題となる。現在の運用方式では事前に書類でのチェックや、対面でのチェックが可能であるが、リモートアクセス型を実現した場合、そのリアルタイム性を損なわない認証方法の検討が必要になる。ID 申請時の審査等はもちろん、利用者の本人確認に関してもシステム化の検討が必要である。また、たとえ本人による利用であったとしても、利用の目的の範囲内であることをアクセスログ等から判断する監査の仕組みを設けることも考えられる。

6 おわりに

本研究では将来的なリモートアクセス型のオーダーメード集計の可能性について、秘密分散・秘密計算技術と汎用集計ソフトウェアを連係させ、実現性と今後の課題の洗い出しを行った。汎用集計ソフトウェアに関しては、現在、統計センターでも利用されている Adam-Report の簡易版 Adam-Lite を用い、同センターから提供されている擬似ミクロデータを利用して評価を行った。

秘密分散・秘密計算技術と汎用集計ソフトウェアの連係に関しては、専用のパイプモジュールを開発して評価を行い、技術的に実現可能である事を確認した。秘密分散・秘密計算技術と汎用集計ソフトウェアを適切に連携させる事により、高いデータの機密性を確保しながら公的データ集計処理を実現することが技術的に可能な事がわかった。しかし、秘密分散・秘密計算技術を用いた集計を行う際には、特に重み付け集計や多次元のクロス表作成時に処理時間が非常に大きくなることも判明し、今後の改善が必要となる。

また、本調査では課題として触れるに留めた秘匿処理もリモートアクセス型のオーダーメード集計を行う上では最重要検討課題の一つである。秘密分散・秘密計算技術を使うことによってデータの保管・処理時の安全性は非常に強固なものにすることができるが、処理そのものの可否については別に適切に判断しなくてはならない。判断のシステム化の検討に加え、その秘密計算上での実現可能性についても今後研究を進めていく必要がある。