Vol.2010-DPS-142 No.2
Vol.2010-CSEC-48 No.2
2010/3/4

2パーティ秘匿回路計算を利用したプライバシー保護データ分析実験報告（1）－CSS2009における行動分析－

五十嵐 大[†1] 千田浩司[†1] 柴田賢介[†1]
山本太郎[†1] 高橋克巳[†1]

2009年10月に開催されたCSS2009では，2パーティ秘匿回路計算システムを用いた行動分析実験が行われた．従来秘匿回路計算は計算コスト等の課題から，実用的な利用報告がほとんどされてこなかった．これに対し本稿では，数百人規模となったCSS2009における上記実験を利用例として報告し，秘匿回路計算及び上記システムの実用性を示す。

Experimental Trials on Privacy-preserving Data Analysis Using 2-party Secure Circuit Evaluation (1) － The Behavior Analysis in CSS2009 －

Dai Ikarashi,[†1] Koji Chida,[†1] Kensuke Shibata,[†1]
Taro Yamamoto[†1] and Katsumi Takahashi[†1]

In CSS2009, convened in October, 2009, an experimental trial on a behavior analysis using a 2-party secure circuit evaluation system was conducted. In the past, there had been very few practical utilization reports of the 2-party secure circuit evaluation, due to its high computational cost mainly. In this paper, we report the above trial as an example of the technique's utilization, and show the utility of the technique and our system above.

†1 NTT 情報流通プラットフォーム研究所
NTT Information Sharing Platform Laboratories

1. はじめに

現在様々なサービスで，年齢・性別，趣味・嗜好，あるいは購買履歴や位置情報といった，個人に関連するデータの利用が進められている．

これらの利用の背景には様々な期待がある．例えば，サービスの利用者にとっては自分により適したサービスを受けられる，サービスの提供者にとっては顧客からより高い満足と収益を得られる，社会全体では生活の改善や社会合理性の追求につながる新しい知見が得られる，等が挙げられる．

しかしこれら個人に関するデータの利用は上記のように利益をもたらす一方で，本来侵害すべきでないプライバシーを侵害してしまう可能性があるという問題を持っている．この問題を解決し，安心かつ安全に個人に関するデータを利用するため，にわかに**プライバシー保護データ分析/マイニング（PPDA/PPDM）**技術の研究が盛んとなってきている．

PPDAは文字通りデータに対してプライバシー保護措置がされるようなデータ分析技術であり，アプローチとして統計的手法を用いるk-匿名法（[文献1)] 等），確率的手法を用いる再構築法（[文献2)] 等），そして暗号学的手法を用いるセキュア計算 ([文献3)] 等) があるが，いずれも現状では社会的受容性等の問題と共に技術的課題も多く，実用は困難で稀であった．

本稿ではこのうちセキュア計算に属する手法である**秘匿回路計算**に関して，筆者らの構築したシステム[4)] を用いたデータ分析実験の実施報告を行う．分析対象は参加者300名を越えた比較的規模の大きい会議である，コンピュータセキュリティシンポジウム2009（CSS2009）[5)] 参加者の属性情報及び行動ログであり，この2種類の情報を結合して分析を行った．属性情報とは後述するシンポジウム参加者への7問のアンケートの回答であり，行動ログとは本実験においてはRFIDシステムにより収集される，RFIDアンテナ付近を通ったRFIDタグのID及びその時刻である．

RFIDシステムにより収集した行動ログや，またその静的な形である位置情報は単独でも“ライフログ”の主要な部分として注目されているが，本実験で行ったような複数の種類の情報の結合は**垂直統合**と呼ばれ，情報量が増加するためより有用な分析が期待される．例えば本実験で抽出できた分析結果として，“学生は[セキュリティと社会] カテゴリのセッションへの参加率が高い”といった，“学生”という属性とセッション参加という行動の相関がある．しかし反面，垂直統合にはデータ保持者が相互に個人に関するデータを提供することの困難さや，“企業所属で，20代後半で，北陸・甲信越出身で，. . . , PPDMセッションに参加し

Vol.2010-DPS-142 No.2
Vol.2010-CSEC-48 No.2
2010/3/4

て, . . . ” などと一人に対するデータが増加した場合匿名性が著しく低下し, 個人の特定の危険性が増してしまうという問題がある.

PPDA の大きな目的のひとつはこれらの問題を解決し**垂直統合データベース上の分析を可能とすること**であり, 本稿の報告は同じシステムを用いた別の実験報告である [文献 6)] と共に, 上記目的を達成するにあたって秘匿回路計算が性能面で現実的に利用しうることを示すものである.

本稿ではまず続く 2 節で PPDA や秘匿回路計算の実用化に関連する研究及び動向を述べ, 3 節で本稿の実験で用いた秘匿回路計算, SCI プロトコルを紹介する. そしてこれを踏まえて 4 節で行動分析実験の条件を説明し, 5 節で実験に用いたシステムについて述べ, 6 節で実施した分析と実験結果について概説し, 最後に 7 節でまとめと今後の課題を述べる.

2. 関連する研究及び動向

本節では PPDA に関する研究や動向をいくつか紹介する.

まずはじめに, 最も実用に近いと言われているのが統計的手法, k-匿名法である. k-匿名法は 2002 年に Sweeney らにより提案された手法[1] であり, 分析対象の個々のデータ（レコード）に対しデータベース管理者が, 一般化または切り落とし等の匿名化処理を行うことで個人の識別を困難とする手法である. k-匿名法の利点として, 残る 2 つの手法と比べ処理が軽く計算コストが現実的であり, さらにデータ分析者は通常の分析を行うだけでよいという点が挙げられる. そのためか k-匿名法に関しては米国において既に製品化の例がある（Privacert Appliance[7]）ほか, 日本においても経済産業省による情報大航海プロジェクト[8] 等において検討がされている.

次に確率的手法を用いる再構築法であるが, こちらは 2000 年に Agrawal らにより提案された[2] 手法である. 再構築法は定められた確率に従って個々のデータを変化させ (撹乱と呼ぶ), ベイズ推定等を用いた再構築と呼ばれる操作により統計値のみを推定する手法である. 再構築法の利点としては以下のようなことが挙げられる. k-匿名法ではデータベース管理者が匿名化処理を行うため個々のデータを知る必要があるが, 再構築法における撹乱は他者のデータを知る必要がないため, データを提供する個人自身がプライバシー保護処理を実行でき, データベース管理者にデータを知られないような利用モデルが実施可能である. また撹乱が非可逆操作であるため, 暗号における鍵を知られるような危険性が存在しない. また計算コストにおいては k-匿名法と比較し得る高速性をもち[9], さらにユーザビリティの観点からも k-匿名法と同じくデータ分析者は通常の分析を行うだけでよいような方式も提案されている[9]. しかし, 再構築法に関してはプライバシー保護の指標が確立されておらず, 利用の大前提となるプライバシーの保証がされていないという課題がある. これに関しては k-匿名法における指標である k-匿名性を拡張した指標, Pk-匿名性を筆者らが提案している[10].

最後に以下 2.1 節で暗号学的手法を用いるセキュア計算について紹介する.

2.1 セキュア計算と秘匿回路計算

セキュア計算は暗号学的な方法を用いて, データは暗号等で保護された状態のまま計算を行い, 結果のみを平文で得る手法である. 他の 2 つのアプローチに比べ,

- 通常の（PPDA でない）分析と等しい分析精度
- 暗号学に基づく強力なデータ保護

を特長とする. k-匿名法, 再構築法におけるプライバシー保護処理は非可逆であり, データが何らかのかたちで変化してしまうため, 非 PPDA と同様の完全な精度をもった分析を行うことは不可能である. これに対しセキュア計算は可逆な処理である暗号を用いるため, 分析結果は非 PPDA と完全に一致する. また k-匿名法や再構築法のプライバシー保護は統計/確率的なものであるため, 小規模のデータに適用するのは困難であるが, セキュア計算は暗号によるビットレベルでのデータ保護であるため小規模データにも適用可能である.

このような特長がある反面, セキュア計算は計算及び通信のコストが非常に高く, また複数の計算者が協力しないと計算ができない[⋆1] という制限がある.

そのため, 基本的なアイデアが 20 年以上前に現れた[12] にも関わらず実装例は数えるほどしかない [文献 14)15)16), 日本では文献 17)18)]. また, 実データを用いるような実用に近い実験例はさらに少ない [文献 15), 日本では文献 17)].

セキュア計算には限定された演算のみを行う特化型の手法と, 任意の論理回路を計算できる汎用型の手法, **秘匿回路計算**が存在する. 例えば文献 15)17)18) は特化型であり, 文献 14) は汎用型の手法である. 特化型は演算を限定する代わりに汎用型よりも計算コストが低く, 上記実験例が両者とも特化型であるのもこの差に起因するものと考えられる.

本稿における行動分析実験は次節で紹介する汎用の秘匿回路計算, SCI プロトコルを採用した秘匿回路計算システム[4] を用いており, 上記の特化型における実験に続いて汎用型も実用的な性能を持ち始めていることを示すものである.

⋆1 理論の段階であるが, 単独で計算できる手法も提案されている[11].

Vol.2010-DPS-142 No.2
Vol.2010-CSEC-48 No.2
2010/3/4

3. 実験で用いた秘匿回路計算：SCI プロトコル

本節では本稿で報告する実験で用いた秘匿回路計算プロトコル[4], SCI プロトコルを紹介する. 本プロトコルは秘匿回路計算プロトコルの中でも 2 人の計算者が合意した計算のみを行うことのできる, 2 **パーティプロトコル**であり, この 2 人が結託しない限り入力データは誰にも漏れることはない.

本プロトコルの基礎は [文献 12)] の難読化回路 (garbled circuit) である. この種のプロトコルは通信回数が定数であるため秘匿回路計算の中でも効率がよいが, 本プロトコルは公開鍵系の処理を一切行わないため, [文献 13)] や [文献 16)] 等で使われる同様の基礎をもつプロトコルと比較しても特に高効率なプロトコルである. 例えば [文献 16)] における最も高速な実装の回路素子当たりの処理速度は約 $200\mu s$ であるが, 本稿の実験で用いた本プロトコルの実装は回路素子当たり約 $11\mu s$[4] と, 20 倍近い差が見られる.

次に本プロトコルの処理を紹介する. 本プロトコルでは *Proxy*, *Server* という 2 パーティの計算者が, データ提供者である *Client*（複数でもよい）の入力を秘匿したまま行う. その処理は 3 つのフェーズに分かれており, 以下のようなものである.

秘匿フェーズ

Client は各入力ビットに対し以下の（1）〜（3）の秘匿処理を行う.

入力：秘密のデータビット $a \in \mathcal{B}$. ただし $\mathcal{B} = \{0,1\}$.

（1） 所定ビット数の乱数 2 つと 1 ビットの乱数 1 つ（$U_0, U_1 \in \mathcal{B}^\kappa, c \in \mathcal{B}$, ただし $\kappa \in \mathbb{N}$ はセキュリティパラメータ）, を生成する.

（2） Client は (U_0, U_1, c) を Proxy に送信する.

（3） Client は $U_a|(a \oplus c)$ を Server に送信する. ただし記号 | はビット列の連結を表す.

このフェーズで Proxy や Server からの応答は不要である.

難読化回路生成フェーズ

次に Proxy は入力を知らずに評価が可能な回路である難読化回路を生成し, Server に送信する. 処理としては以下の回路素子 1 個を実現する次の処理（1）〜（2）を再帰的に行う.

入力：$(U_0, U_1, c), (V_0, V_1, d) \in \mathcal{B}^\kappa \times \mathcal{B}^\kappa \times \mathcal{B}$

（1） 出力用に, 所定ビット数の乱数 2 つと 1 ビットの乱数 1 つ（$W_0, W_1 \in \mathcal{B}^\kappa, e \in \mathcal{B}$）を生成する.

（2） Server に難読化真理値表を送信する. すなわち任意のビット組 $(c', d') \in \mathcal{B}^2$ に対し

$$\begin{cases} (W_{g(c',d')}|e \oplus g(c',d')) \oplus H(U_{c'}|d') \oplus H(V_{d'}|c') & \text{最終段以外} \\ (0|g(c',d') \oplus H(U_{c'}|d') \oplus H(V_{d'}|c') & \text{最終段} \end{cases}$$

を送信する. ただし $H : \{0,1\}^{\kappa+1} \to \{0,1\}^{\kappa+1}$ はハッシュ関数を表す.

回路初段での入力は Client からの入力そのものである. 2 段目以降は前段で生成した (W_0, W_1, e) を再帰的に入力とする. 上記処理を注意して見ると, 難読化回路の生成自体は通信を必要としないことが分かる. このため Proxy-Server 間の通信回数はただ 1 回のみで十分ある.

難読化回路評価フェーズ

最後に Server が難読化回路を評価し, 計算結果を得る. ここでも以下の回路素子 1 個に対する処理（1）を再帰的に行う.

入力：$U|a', V|b' \in \mathcal{B}^{\kappa+1}$（ただし $U, V \in \mathcal{B}^\kappa, a', b' \in \mathcal{B}$）と, 難読化真理値表すなわち, 任意の $(c', d') \in \mathcal{B}^2$ に対し $w_{c',d'}$（ただし各 $w_{c',d'} \in \mathcal{B}^{\kappa+1}$）.

（1） $w_{a',b'} \oplus H(U|b') \oplus H(V|a')$ を計算する.

上記各処理において最も重い処理は Proxy の難読化回路生成であり, 回路素子 1 個につきハッシュ関数 8 回の計算が必要と, Server の難読化回路評価のほぼ 4 倍の処理となる. Client の処理は自身の入力データの分のみで, Proxy や Server が行う回路全体の処理に比べると一般に微々たるものである.

4. CSS2009 行動分析実験の概要

本節では CSS2009 で行われた, 秘匿回路計算を用いた行動分析実験について述べる.

本実験の目的

本実験の目的は CSS 運営側としては CSS2009 の統計情報を得ることがひとつにあるが, 筆者らの大きな目的は, 秘匿回路計算を用いたシステムが, CSS2009 という比較的大きな規模の現実の会議において実験ではあるが真に動作することを検証することである.

収集したデータ

本実験で分析対象とするために収集したデータは以下の 2 種類である.

（1） アンケートによる, CSS2009 参加者の**属性情報**

（2） RFID システムにより収集される, CSS2009 参加者の**行動ログ**

属性情報とは参加者の例えば性別や年齢などの情報であり, 本実験では表 1 記載の 7 種類の属性である.

Vol.2010-DPS-142 No.2
Vol.2010-CSEC-48 No.2
2010/3/4

表 1 本実験で実施したアンケートの設問

設問	選択肢数
所属区分	5
年齢	8
出身地	10
血液型	4
今熱いと思う研究テーマ	2（×23 テーマ）
ファイル交換ソフトを利用していますか?	3
好きなお酒の種類	6

行動ログは本実験では “どのセッションに入退出したか” という情報である．これらの情報を統合すれば, 例えば “暗号セッションに来た参加者は 20 代が多い” などといった, 2 種類の情報の組み合わせによりはじめて導かれる結果を得ることが期待される．

属性情報の収集方法

属性情報は会場に受付をはじめとして設置した 4 台の PC 端末を介して, 参加者にブラウザベースのアンケートに答えてもらうことで収集した．

行動ログの収集方法

行動ログについては約 3m 程度の距離からタグを検知することのできる, UHF 帯 RFID システムを用いて収集した．CSS2009 の 6 つ全てのセッション会場出入り口に RFID アンテナを配置するとともに, 実験への協力に同意する参加者に RFID タグを配布し, 名札とともに身につけてもらった．このタグを会場のアンテナが検知し, 検知時刻及びタグ ID を記録するという方法である．

本実験におけるプライバシー保護のモデル

前述の通り本実験で用いた SCI プロトコル[4] においては, 2 人の計算者が合意した計算のみを行うことができ, この 2 人が結託しない限り入力データが各データ提供者本人以外に漏れることはない．本稿ではプロトコルへの入力データをシステム上の他の入力と区別するため, **秘匿データ**と表記する．

本実験で秘匿データに該当するのは属性情報である．属性に関するアンケート回答はサーバー送信以前に後述する秘匿処理により秘匿データとなるため, 誰にも漏れることはない．行動ログに関しては秘匿データとはならないが, RFID タグを参加者自身に無作為に選んでもらうことで RFID と参加者個人が結びつかないようにし, 匿名性を確保した．

また 2 人の計算者としては, CSS2009 実行委員長の高橋克巳及び CSS2009 プログラム委員長の西垣正勝教授に秘匿データの暗号鍵の管理を委任することで 2 人の結託なしに秘匿データが漏れないようにした．

5. 本実験に用いたシステム

本実験においては以下の 3 つのシステムを併せて用いた．ひとつは SCI プロトコルの実装である**セキュア表計算システム**[4], 次にアンケート回答を秘匿する秘密アンケートシステム, 最後に行動ログを収集する, RFID システムである．以下でそれぞれのシステムに関して紹介する．なお, 3 つのうちセキュア表計算システム, 秘密アンケートシステムは [文献 6)] でも用いられている．

5.1 セキュア表計算システム

本節では文献 4) の秘匿回路計算システム（以降セキュア表計算システムと呼ぶ）を紹介する．

セキュア表計算システムは Microsoft Excel[⋆2] をフロントエンドとする秘匿回路計算システムである．秘匿回路計算エンジンは前述の SCI プロトコルを採用しており, 現在

最大値/最小値/中央値/クロス集計/平均値/最頻値/分散/加算/減算/乗算/平方算

の 11 の演算が実装されている．

本システムにおけるデータの流れは図 1 のようになっており, **分析クライアント**がデータ管理者となって**データ提供者**たちから暗号化データを収集し, **秘匿回路計算サーバ**に計算を依頼して計算結果を得る, という想定である．分析クライアントは以下の手順でデータ収集及び分析を行うことができる．

分析の手順

(1) 分析クライアントが秘匿用 Excel シートを作成し, *Server*, *Proxy* の公開鍵と共に各データ提供者に配布する．
(2) 各データ提供者は専用の Excel アドイン（**秘匿回路計算アドイン**）によりデータを秘匿処理し, 分析クライアントに送信する．
(3) 分析クライアントは Excel 操作等によりデータを集約し, 専用 Excel アドインを用いて *Controller* に計算リクエストを送信する．
(4) 秘匿回路計算サーバで計算が行われ, 分析クライアントは結果を受信する．

上記 Controller, Proxy, Server は秘匿回路計算サーバの構成要素であり, Proxy, Server は

⋆2 Microsoft, Excel は共には米国 Microsoft Corporation の米国およびその他の国における登録商標または商標です．

Vol.2010-DPS-142 No.2
Vol.2010-CSEC-48 No.2
2010/3/4

3 節のそれぞれに相当し実際の秘匿回路計算を担当する．Controller は分析クライアントとのインタフェースであり，SCI プロトコルから見れば単なる通信路である．また分析クライアントも同様にプロトコルから見れば通信路に相当する．

3 節の Client に相当するデータ提供者がセキュア表計算上で行う秘匿処理は，Proxy, Server へ到達する通信路のセキュリティを確保するため，3 節の秘匿処理を施した入力にさらに Proxy, Server の公開鍵で暗号化する，というものとなる．なお全データを公開鍵暗号化するのは計算コストがかかるため，本システムでは 1024bitRSA と共通鍵暗号 Camellia[19]（128bit, CBC モード）とのハイブリッド暗号を用いている．

この通信路用の暗号化のため，秘匿回路計算サーバの処理にも対応する復号が含まれ，以下のようになる．

秘匿回路計算サーバの処理

（1） Controller は Server, Proxy に，分析クライアントのリクエストと暗号化秘匿データを送信する．

（2） Server, Proxy は通信路用暗号の復号により秘匿データを得，両者間で秘匿回路計算を行う．

（3） Server, Proxy のうち，計算結果を受け取るパーティである Server が Controller を介して分析クライアントに計算結果を送信する．

なお，秘匿時/通信路用暗号復号時の RSA 及び Camellia 暗号化/復号，そしてハッシュとして用いている Camellia（128bit, ECB モード）暗号化はいずれも OpenSSL 0.9.8i[22] を用いた．

5.2 秘密アンケートシステム

秘密アンケートシステムは属性情報の収集に用いたシステムである．本システムは図 2 のようにブラウザ上でユーザにアンケートを行い，セキュア表計算システム互換の秘匿処理が施された回答結果を Web サーバに蓄積するシステムであり，以下の構成要素から成っている．

（1） 回答結果を秘匿する JavaScript

（2） 回答結果を保存する CGI

回答結果の秘匿が CGI ではなく JavaScript で行われるのは，データ提供者に属しないマシンである Web サーバに到達する前にデータを秘匿するためである．またアンケートには RFID タグ記載の 16 進 6 桁の “RFID パスワード” 欄が存在する（図 2 下部）．このパスワードは後述のタグ ID と 1 対 1 対応しており，分析時に行動ログとのデータ結合に用いる

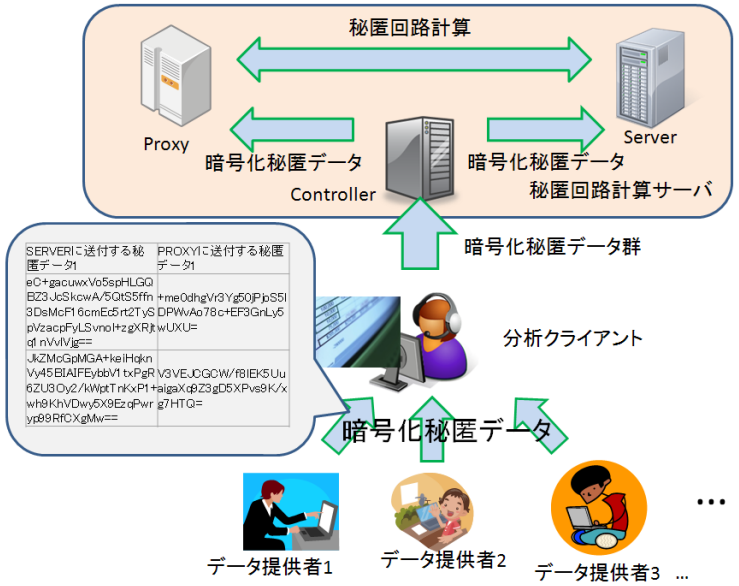

図 1 セキュア表計算システム

ことができる．本パスワードは手入力のため，故意又はミスにより誤った ID をアンケート回答の際に入力した際に結果が不正になる可能性が十分小さく，かつ入力の手間も大きすぎない程度のサイズとして，上記の通り 16 進 6 桁の乱数とした．

なお上記 JavaScript においては Camellia による暗号化には [文献 20)]，RSA による暗号化には [文献 21)] のライブラリを用いた．

5.3 RFID システム

RFID システムは行動ログの収集に用いたシステムである．RFID はアンテナとタグ間で無線通信を行うことで，アンテナ付近のタグの存在を検知することができる．このアンテナをいくつか設置することでタグ装着者の位置情報，そしてこれを連続的に取得することで行動ログを得ることができる．位置情報や行動ログはライフログの重要な部分として近年注目されている情報であり，この観点でも本実験は意味を有する．

本システムは UHF 帯 RFID を用いており，約 3m の距離からタグを検知できるためデータ提供者にカードをかざすなどの行為を求めることなくアンテナ前方の通過を検知するこ

Vol.2010-DPS-142 No.2
Vol.2010-CSEC-48 No.2
2010/3/4

CSS2009行動分析実験アンケート

所属区分	学生
年齢	19才以下
出身地	北海道
血液型	A型
今熱いと思う研究テーマ（複数回答可）	□暗号・評価 □署名・暗号プロトコル □情報ハイディング □ネットワーク監視・追跡 □コンピュータウィルス □Web・メールセキュリティ □アクセス制御 □認証・バイオメトリックス □セキュリティ設計・実装 □OS・仮想化 □ハードウェア □ユビキタスセキュリティ □電子商取引 □コンテンツ保護 □ソフトウェア保護 □リスク分析・セキュリティポリシー □セキュリティ評価・監査 □個人情報・プライバシー保護 □フォレンジクス □セキュリティ教育・法律 □SPT(心理学とトラスト) □マルウェア対策 □その他
ファイル交換ソフトを利用していますか？	利用したことはない
好きなお酒の種類	ビール
RFIDカードのPWを入力してください	（英字は大文字で入力お願い致します）

図 2 CSS2009 行動分析の際の秘密アンケート

とができる．参加者が自身の被検知を確認し実感できるよう, 検知されたタグ ID は会場に設置したモニターにリアルタイム/グラフィカルに表示した（図 3 左）．この目的でタグに存在する 128bit の書き込み可能領域にあらかじめ 1〜400 の ID を書き込み, タグ表面にも当 ID を記載した．

なおハードウェアはオムロン社のリーダライタ V750-BB50C04-JP 及びアンテナ V750-HS01CA-JP(図 3 右下), タグ V750-D22M01-IM（図 3 右上）を用いた．ソフトウェアに関してはオムロンソフトウェア社による評価ライブラリを利用して, C#により制御ソフトウェア（図 3 左）を実装した．

6. 本実験で実施した分析と結果

本節では本実験で行った分析と, 実験結果について述べる．

データ収集結果

まず収集できたデータ量については表 2 の通りである[5]．重複を除くタグの検知枚数から, CSS2009 参加者の半数以上が実験への協力をしたこと, またアンケート回答数からは参加者の 1/4 以上がアンケートに回答したことが分かる．

分　　析

分析は Microsoft Excel 上に秘密アンケートシステムによる属性情報及び RFID システムによる行動ログをすべて集約し, 前述のタグ ID 及び RFID パスワードをキーにしてデータ結合を行った．実施した分析は以下の 3 つである．なお実験結果の詳細・グラフは [文献 5)] を参照されたい．ここでは例として分析 3 の一例のみを図 4 に掲載する．

図 3 左：RFID 制御ソフトウェア表示画面, 右上：RFID タグ（V750-D22M01-IM），右下：RFID アンテナ（V750-HS01CA-JP）

表 2 本実験で収集できたデータ

会期中のタグ検知数 (のべ数)	2339 回
同上 (重複除く)	215 枚
アンケート回答数	97 名

※ CSS2009 参加者：362 名

(1) 分析 1：各セッションの参加数

CSS2009 各セッションの重複を除いたタグ検知数を計測し，それぞれの参加人数の傾向を探った．CSS2009 会期中に各セッションで検知された RFID タグ数（重複除く）をグラフにした．

(2) 分析 2：熱いテーマ Top10

アンケート項目 “熱いと思うテーマは？(複数回答可)” から，CSS2009 参加者の興味を探る．所属区分別（アンケート回答者全体, 学生, 教員, 企業, 独法）に, アンケート設問 “今熱いと思う研究テーマ（複数回答）” 中の各研究テーマについて，チェックした人の割合を “熱い率” として Top10 をグラフにした．

Vol.2010-DPS-142 No.2
Vol.2010-CSEC-48 No.2
2010/3/4

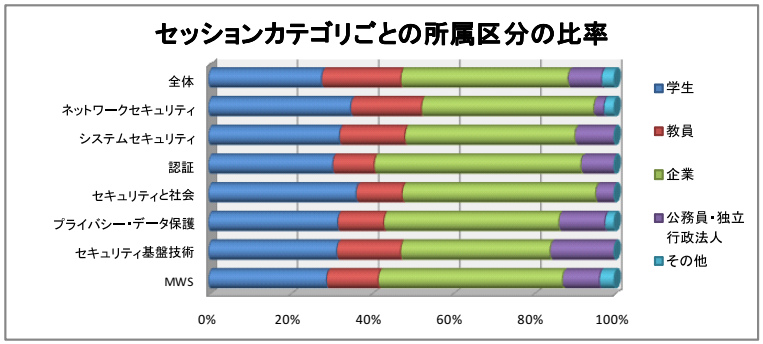

図 4 分析結果の例

(3) 分析 3：セッションとアンケート回答の関係
行動ログとアンケート結果をクロス集計し，セッションと属性の関係を探る．各セッションを 7 つのカテゴリに分類し，アンケート回答者全体及び各カテゴリのセッションで検知された人の，アンケート回答結果の比率をグラフにした．またこの各カテゴリの RFID 検知ログとアンケートを分析した結果，いくつかの相関が見られたのでその結果も表記した．

分析 1 については秘匿データに該当しない，行動ログのみからの分析であり，秘匿回路計算を利用していない．分析 2 は秘匿データであるアンケートからの分析であり，クロス集計演算の秘匿回路計算を各研究テーマについて，すなわち 23 回実行した．分析 3 は行動ログ及びアンケート両方を用いた分析であり，垂直統合データの分析にあたる．ここでは行動ログをセッション時間ごとに対応する時間帯でフィルタし，フィルタ後のデータに対して秘匿回路計算を行った．演算としてはこちらもクロス集計であり，7 カテゴリに分けたセッションカテゴリごとに 6 つの設問，すなわち 42 回の秘匿回路計算を行った．

分析 2, 3 ではセキュア表計算システム上で，計 65 回の秘匿回路計算を実行した．セキュア表計算システムの性能は [文献 4)] で既に評価したため特に計測は行わなかったものの，1 回ごとの演算は数秒で終了し，Microsoft Excel 上の操作も含めても 30 分以内に終了した．

7. おわりに

本稿では CSS2009 で行われた，秘匿回路計算を用いた行動分析実験について報告した．この行動分析実験はアンケートデータと RFID システムによる行動ログとを垂直統合して統計分析を行うものである．データの垂直統合分析はより有用な情報を得られる期待がある一方でプライバシー保護の問題があるという，秘匿回路計算のようなプライバシー保護データ分析技術の最も重要な応用例のひとつであり，本実験では参加者の半数以上から行動ログ，そして 1/4 以上から属性情報の提供を得て統合分析を実現した．

本実験では [文献 4)] による秘匿回路計算，SCI プロトコルの実装であるセキュア表計算システムを実際に動作させ分析を行った．この分析では 65 回の秘匿回路計算を行い，フロントエンド上の操作まで含め 30 分以内に計算が終了した．このことから従来計算コストのために現実的でないと思われていた秘匿回路計算が，上記のような重要な応用例に対して既に実用に耐えうる性能に達し始めていると言える．

今後はより処理性能を高めることで，ライフログ等のより高度に蓄積されたデータへの適用を目指す予定である．例えば近年は CPU のマルチコア化，GPU による汎用並列計算環境の出現等により並列計算資源が豊富になっているため，これらの利用が考えられる．また [文献 23)] の技法の適用等も数倍の高性能化を見込むことができる．

さらに性能面のみでなく，[文献 6)] で行ったような，秘匿回路計算の採用によるユーザのデータ提供量意思及び収集データ量の変化等の検証，そして今回の実験では実装済演算の制限から RFID システムによる行動ログを秘匿して演算することができなかったが，演算の充実によりこのようなデータの秘匿回路計算にも対応するなど，適用範囲の拡充等も行っていく予定である．

謝辞 タグ装着及びアンケート回答に協力下さった CSS2009 の皆様，実験に不可欠な秘密鍵管理を請け負って下さった CSS2009 プログラム委員長の静岡大学西垣正勝先生，また秘密アンケートシステム，通称 “関野システム” の作成を手がけて下さった中央大学の関野智啓さんに感謝いたします．ありがとうございました．

参 考 文 献

1) L. Sweeney. k-anonymity: a model for protecting privacy. *Int'l Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol.10, Issue 5, pp.557-570, 2002.
2) R. Agrawal and R. Srikant. Privacy-preserving data mining. *Proc. of the 2000 ACM SIGMOD Intl. Conf. on Management of Data,* 2000.
3) Y. Lindell and B. Pinkas. Privacy Preserving Data Mining. *CRYPTO, Vol. 1880 of LNCS*, Springer, pp. 36-54, 2000.
4) 柴田 賢介, 千田 浩司, 五十嵐 大, 山本 太郎, 高橋 克巳. 表計算ソフトをフロントエン

Vol.2010-DPS-142 No.2
Vol.2010-CSEC-48 No.2
2010/3/4

ドとした委託型 2 パーティ秘匿回路計算システム. *CSS2009,* 2009.
5) コンピュータセキュリティシンポジウム 2009. http://www.iwsec.org/css/2009/.
6) 柴田 賢介, 千田 浩司, 山本 太郎, 高橋 克巳, 金井 敦. 2 パーティ秘匿回路計算を利用したプライバシー保護データ分析実験報告 (2) -大学生の成績と生活実態との相関分析-. 第 142 回 DPS・第 48 回 CSEC 合同研究発表会, 2010.
7) Privacert. Privacert Appliance. http://www.privacert.com/appliance/index.html.
8) 経済産業省, 情報大航海プロジェクト, http://www.igvpj.jp/index/.
9) 永井 彰, 五十嵐 大, 濱田 浩気, 松林 達史. クロネッカー積を含む行列積演算の最適化による効率的なプライバシー保護データ公開技術. *SCIS2010,* 2010.
10) 五十嵐大, 千田浩司, 高橋克巳. k-匿名性の確率的指標への拡張とその適用例. *CSS2009*, 2009.
11) C. Gentry. Fully Homomorphic Encryption Using Ideal Lattices. *STOC2009*, 2009.
12) A. C. Yao. How to Generate and Exchange Secrets. *Proceedings 27th Symposium on Foundations of Computer Science (FOCS)*, EEE, 1986, pp. 162-167.
13) D. Malkhi, N. Nisan, B. Pinkas, Y. Sella. Fairplay - a secure two-party computation system. *Proc. of 13th USENIX Security Symposium (2004)* , 2004.
14) A. Ben-David, N. Nisan, B. Pinkas. FairplayMP: a system for secure multi-party computation. *Computer and Communications Security (CCS) 2008* , pp. 257-266. ACM, New York, 2008.
15) P. Bogetoft, D.L. Christensen, I. Damgård, M. Geisler, T. Jakobsen, M. Kroigaard, J.D. Nielsen, J.B. Nielsen, K. Nielsen, J. Pagter, M. Schwartzbach, T. Toft. Secure multiparty computation goes live. *FC 2009,* 2009.
16) B. Pinkas, T. Schneider, N.P. Smart, S.C. Williams. Secure Two-Party Computation is Practical. *ASIACRYPT2009,* 2009.
17) 独立行政法人情報通信研究機構 (NICT). 報道発表, スピア型サイバー攻撃判定システム開発のための共同実証実験を開始, http://www2.nict.go.jp/pub/whatsnew/press/h19/080303/080303_1.html.
18) 佐藤 夏樹, 篠原 靖志. プライバシーを保護した需要データ収集・共用方式の開発（その１）－需要データ収集・需要特性算出方式－. http://jglobal.jst.go.jp/public/20090422/200902267615822378, 2008.
19) NTT 情報流通プラットフォーム研究所. Camellia 紹介. http://info.isl.ntt.co.jp/crypt/camellia/intro.html.
20) 小山 浩之. 100% Pure JavaScript Camellia. http://alpha.mixi.co.jp/blog/?author=12.
21) Tom Wu. BigIntegers and RSA in JavaScript. http://www-cs-students.stanford.edu/~tjw/jsbn/.
22) OpenSSl. www.openssl.org/.
23) V. Kolesnikov, T. Schneider. Improved garbled circuit: Free XOR gates and applications. *ICALP 2008, Part II.*, LNCS, vol. 5126, pp. 486-498. Springer, Heidelberg, 2008.