

明治大学総合数理学部

2023 年度

卒業研究

ユーザ属性セキュリティ意識と国の違いによる多要素認証方式
の志向性の調査

学位請求者 先端メディアサイエンス学科

小池蒼葉

目次

第 1 章	はじめに	1
第 2 章	先行研究	2
第 3 章	多要素認証の利用調査	3
3.1	Meiji Mail 多要素認証の種類	3
3.2	方法	3
3.3	SeBIS	3
第 4 章	Meiji Mail のユーザを対象にした調査	5
4.1	概要	5
4.2	セキュリティ意識と属する学部に関連について	5
4.3	認証方式の人気度・セキュリティ意識との関連	5
4.4	考察	6
第 5 章	国の異なる全ての人を対象にした調査	10
5.1	概要	10
5.2	国の違いとセキュリティ意識の違い	10
5.3	認証方法の人気度・セキュリティ意識との関連	10
5.4	考察	11
第 6 章	おわりに	16
参考文献		17
付録 A	k 匿名化ツールである Nithin Prabhu の改良と評価	18
A.1	はじめに	18
A.2	先行研究	20
A.3	提案システム	20
A.4	おわりに	22
参考文献		23

第1章

はじめに

今日、インターネット上のサービスのアカウントの認証方式は、セキュリティ強化を目的として2つ以上の認証方式を組み合わせた多要素認証が主流である。その要素には、パスワードなどの知識情報、携帯電話などの所持情報、指紋や静脈などの生体情報が挙げられる。中でも多く採用されているのが知識情報である。これはパスワードと何かしらを組み合わせた二要素認証である。明治大学のメールシステムである Meiji Mail も学外からのサインインの際の認証方式として採用している。[1] 第一要素として ID とパスワード、第二要素としてモバイルアプリ、または、SMS で送信される認証コード、電話メッセージを使用することができる。これらの第二要素は、所持情報である。

しかし、この多要素認証の脆弱性を悪用してパスワードを初期化する PRMitM (Password Reset Man in the Middle) 攻撃が Gelernter らにより提案されている。[4] PRMitM 攻撃は、中間者攻撃であり、アカウント登録とパスワードリセットの手順の類似性を利用してユーザのパスワードを初期化する。[5] 一方、モバイルアプリでの認証は SMS コードを使用しないため中間者攻撃の恐れが無い。従って、Meiji Mail の3つの認証方式の中ではアプリ認証を推奨すべきであると考えるが、現状は SMS 認証の利用が多い。また、人の特性の中でも居住国とセキュリティ意識の差や多要素認証方式の関連性があるのではないかと推測する。

そこで、Meiji Mail のサインイン方式の利用の実態と、人の特性やセキュリティ意識との関連性について調査し、利用の理由を明らかにする。アプリ認証を推奨するための必要な条件を探る。

本研究の目的は、多要素認証方式の選択に人の属性とセキュリティ意識がどう関連するかを考察することである。Meiji Mail における調査に加えて、任意のインターネットサービスの認証方式について国の違いとセキュリティ意識、認証方式の人気度の関連性を調査する。

第 2 章

先行研究

2020 年に野口らは, [2] において “レベル分けによる多要素認証の要否の実現と認証連携の拡張” を発表した. 統合認証基盤における認証の要を Active Directory Federation System から Azure Active Directory へ更新し, 多要素認証をもともと使っていた Office365 の他にも Shibboleth 環境でも可能とし, 対象サービス等によりレベル分けされた多要素認証の要否の制御も可能とする多要素認証システムを構築した. また, 同時にネットワーク接続時の認証を認証 s 連携の対象とする拡張おも行った. 結果としては, レベル分け多要素認証を実現でき, 有線ネットワーク接続時の認証をシングルサインオンの対象とすることに成功した. 多要素認証方式の選択状況において, SMS での認証が圧倒的に多かったことを示した.

2023 年に Lyastani らは, [3] において “A Systematic Study of the Consistency of Two-Factor Authentication User Journeys on Top-Ranked Websites” を発表した. トップクラスのウェブサイトにおける二要素認証のユーザージャーニーの外部的, 機能的一貫性に関する体系的研究である. 対象のウェブサイトは最低限のデザイン面は一貫して実装されているが, 第二要因の設定や使用については, 多種多様なデザインパターンを示しているとし, 2FA 実装者のためのより一般的な UX ガイドラインを提唱した. また, 2FA のユーザージャーニーに関する新たな研究課題を提示した.

第3章

多要素認証の利用調査

3.1 Meiji Mail 多要素認証の種類

Meiji Mail は学外からのサインインに多要素認証を採用している。第一要素として ID とパスワード、第二要素としてモバイルアプリ、SMS で送信される認証コード、または電話メッセージがある。電話メッセージでの認証は主流ではないため、それ以外の 2 つが主に使われている。SMS コードとモバイルアプリ (Microsoft Authenticator) の認証の使用感の違いを調べるために、複数回自身でサインインをし時間を計測した。結果、メールアプリ (Out look) を使用の場合はアプリの方がより速くサインインをすることができ、ブラウザ (safari) を使用の場合は SMS コードの方がより速いことが分かった。タイムの差はいずれの場合も約 10 秒ほどであった。

3.2 方法

本研究は Google Form を用いたアンケートによって行った。アンケート項目は、性別や年齢等の基本情報とセキュリティ志向度指標 SeBIS の質問と多要素認証の認証方法についての質問で構成されている。実験は Meiji Mail のアカウントを持つユーザ間での比較を対象にしたものと、国外の人を対象にしたものの 2 種類である。Meiji Mail のアカウントを持つユーザには、多要素認証の認証方法の質問にて Meiji Mail の認証方式についてのものと、それ以外のスマートフォンを使用した多要素認証について尋ねた。質問項目について、Meiji Mail のものについては採用している認証方式やその選択理由、それ以外のものについては最も利用頻度の高い認証方式や中間者攻撃の認知等である。また、国外の被験者にはスマートフォンを使用した多要素認証についての質問に限定した。質問項目について、最も利用頻度の高い認証方式や中間者攻撃の認知等である。

3.3 SeBIS

SeBIS は Serge Egelman らが提案したセキュリティ志向度の指標である。この尺度は、一般的なセキュリティに関するアドバイスに対するユーザが従う程度の治療値である。「ほとんどのユーザに適用可能なセキュリティ行動」と「広く受け入れられているセキュリティ行動」に大別し質問で構成されている。それぞれの質問は肯定的な表現と否定的な表現を混ぜながら作られており、5 段階のリッカート尺度 ([強くそう思わない] から「強くそう思う」) で回答される。表 3.1 で示す Meiji Mail ユーザを対象に行った実験で用いた質問項目は、筐らによって [6] 和訳されたものである。また、表 3.2 で示す国外の人を対象にした実験で用いた質問項目は、

表 3.1 SeBIS 質問項目 (日本語)

番号	質問内容
1	コンピュータを長時間放置したとき、自動的にロックするような設定にしている。
2	ノートパソコンやタブレットのロックを解除するとき、パスワード/パスコードを使っている。
3	コンピュータから離れる時、手動で画面をロックする。
4	携帯電話のロックを解除するために PIN またはパスコードを使用する。
5	必要があるときしかパスワードを変更しない。
6	使っているアカウントごとに違うパスワードを使っている。
7	新しいオンラインアカウントを作るとき、必要最低限の文字数を超えるパスワードを設定する。(8 文字以上なら、9 文字以上で設定)
8	必要が無い場合は、パスワードに特殊文字 (¥や*) を含めない。
9	リンクが送られてきたとき、どこにつながるか確認しないでクリックする。
10	どのサイトに訪れたかを URL ではなくサイトの外観と雰囲気かで判断している。
11	安全な通信が確認することなくウェブサイトへ情報を提出する。(例: SSL, "https://", ロックアイコン)
12	リンクをクリックする前に、マウスアイコンをリンクに乗せ訪れる URL を確認する。
13	セキュリティ上の問題が発見されても誰かが直すだろうからそのまま使い続ける。
14	ソフトウェアのアップデートについてのメッセージが表示されたらすぐにインストールする。
15	使用しているプログラムが最新であることを確認するようにしている。
16	自分のアンチウイルスソフトウェアが定期的に更新されていることを確認する。

表 3.2 SeBIS 質問項目 (英語)

Number	Questions
1	Whenever I step away from my computer, I lock the screen.
2	I use a password/passcode to unlock my laptop or tablet.
3	I manually lock the screen when leaving the computer.
4	I use a PIN or passcode to unlock the phone.
5	Once I create a password, I tend to never change it.
6	I use a different password for each account I use.
7	When creating a new online account, set a password that exceeds the minimum number of characters required. (If it is more than 8 characters, use 9 or more.)
8	Do not include special characters (! or *) in passwords unless necessary.
9	When a link is sent to me, I click on it without checking to see where it leads.
10	I judge which site was visited by the appearance and atmosphere of the site, not the URL.
11	I submit information to a website without verifying that the data communication is secure. (Ex. SSL, "https://", a lock icon)
12	When browsing websites, I frequently mouseover links to see where they go, before clicking them.
13	Even if a security problem is discovered, someone will fix it, so I continue to use it.
14	I apply software updates as soon as my computer prompts me.
15	I make sure that the programs I'm using are up-to-date.
16	I check that my anti-virus software is regularly updated.

それらをそれらを再度自身で英訳したものである。

第4章

Meiji Mail のユーザを対象にした調査

4.1 概要

本調査は Meiji Mail のアカウントを持つ 57 名のユーザを対象に行った。ユーザの基本情報を表 4.1 に示す。アンケート項目は基本情報に加えて、SeBIS(セキュリティ志向度指標)、Meiji mail の多要素認証、Meiji mail 以外のスマートフォンを使用した多要素認証の各々について計 4 種類から成る。人の特性やセキュリティ意識と多要素認証方式の関係について明らかにする。

4.2 セキュリティ意識と属する学部に関連について

取り扱う学業柄、総合数理学部の学生はセキュリティ意識が他の学部生と比べて高いのではないかと考えた。これについて、総合数理学部とその他の学部別の SeBIS の合計得点の箱ひげ図を表 4.1 に示す。この箱ひげ図から、総合数理学部のものの方がデータの範囲が広いことや外れ値があることが分かる。総合数理学部は最大値が 67、最小値が 34 で 2 つの外れ値をデータ範囲が 33 であるのに対し、その他学部の最大値は 58、最小値は 38 でデータ範囲は 20 である。よって、総合数理学部の方がその他学部よりも 13 点データ範囲が広い。一方、平均得点、中央値はそれぞれ総合数理学部が 49.7、45.5 でその他学部が 47.5、46 とあまり違いが見られない。これらのことから、総合数理学部には著しくセキュリティ意識が高い人が一定数いるが、それらを除くと他学部の人とさほどセキュリティ意識に差は見られないことが分かる。

4.3 認証方式の人気度・セキュリティ意識との関連

セキュリティ意識が高い人は認証方法でモバイルアプリを選択するのではないかと予測した。背景には、第一章で述べたように SMS コードによる認証は中間者攻撃の恐れがあることから安全性が無いことが提起されているため、その存在を知っている人はアプリを選択すると考えられたことがある。Meiji Mail の認証方式の人気割合を図 4.2 で示す。SMS コードが 40 人で全体の 70.2%、モバイルアプリが 17 人で 29.8% と SMS コードがより人気であった。また、図 4.3 には認証方法ごとの SeBIS の得点分布を箱ひげ図で示す。このグラフから、モバイルアプリのものの方が全体的なデータの値が高いことが分かる。SMS コードのもの外れ値を除く最大値は 58 で最小値は 34、平均値は 47 であるのに対し、モバイルアプリのものは最大値が 77 で最小値が 41、平均値が 52 とより高得点を記録している。このことから、モバイルアプリ使用者の方が SMS コード使用者よりも高いセキュリティ意識を持つ人が多いことがわかった。

表 4.1 Meiji Mail ユーザ実験基本情報

		人数(名)	割合(%)
性別	男性	17	29.8
	女性	40	70.2
学年	1年	8	14.0
	2年	11	19.3
	3年	10	17.5
	4年	27	47.4
	教授	1	1.8
学部	総合数理	26	45.6
	国際日本	20	35.1
	情報コミュニケーション	4	7.0
	農	1	1.8
	文	2	3.5
	政治経済	1	1.8
	商	3	5.3
留学生		1	1.8

表 4.2 認証方法に SMS コードを選んだ理由

理由	数
馴染みがあるから	16
他の方法(があるの)を知らなかったから	10
アプリが使用できなかったから	2
特に覚えていない	11

4.4 考察

本実験結果より、学部とセキュリティ意識の関連は見られなかったが、認証方式の主流が SMS コードである中で、セキュリティ意識が高い人ほどモバイルアプリでの認証を傾向がわかった。しかし、中間者攻撃の存在を知っている人が必ずしもモバイルアプリでの認証を選択しているわけではないため、また別の理由からアプリを選択している可能性が考えられる。また、各認証方法を初期設定時に選んだ理由を表 4.2 と表 4.3 に示す。表 4.2 から SMS コードの選択理由は“馴染みがある”ことや“他の方法を知らない”ことが多いことが分かる。また、表 4.3 からモバイルアプリの選択理由は“他の方法を知らない”ことが多く、“馴染みがある”ことはあまり多くないことが分かる。これらのことから、SMS コードでの認証が主流であることは慣れ親しんでいることが大きな要因となり、モバイルアプリでの認証を試すことを妨げているのではないかと考えられる。よって、モバイルアプリでの認証を増やすためには、様々な人気ウェブサイトやツールがモバイルアプリでの認証を推奨することが必要であると考えられる。

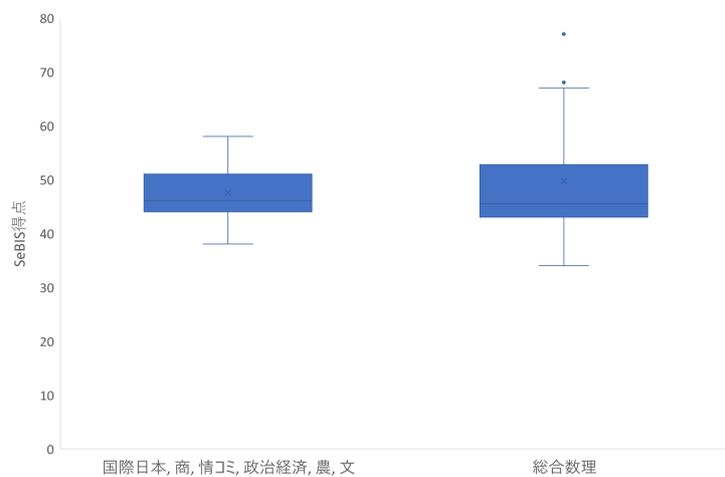


図 4.1 学部別の SeBIS 得点分布箱ひげ図

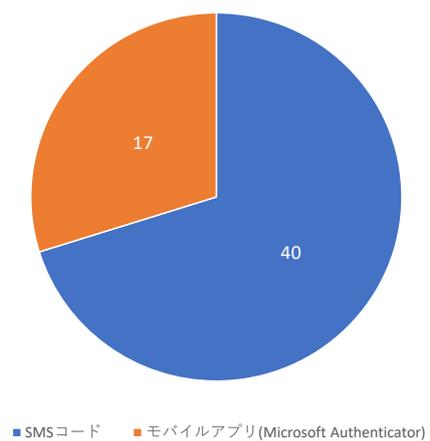


図 4.2 認証方法の割合

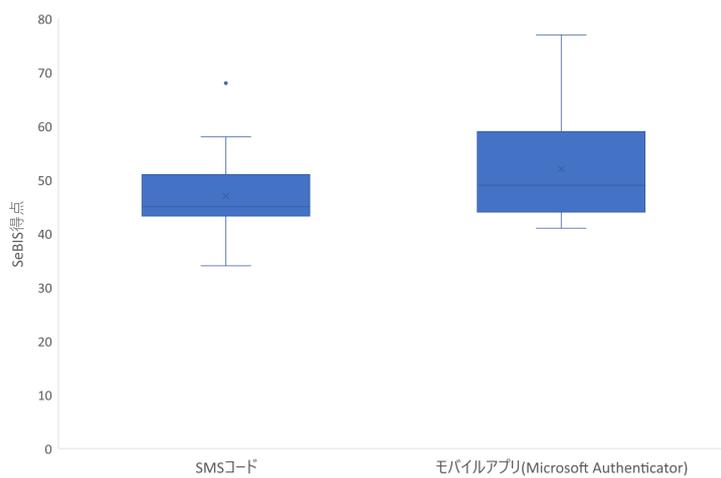


図 4.3 認証方法別の SeBIS 得点分布箱ひげ図

表 4.3 認証方法にモバイルアプリを選んだ理由

理由	数
他の方法 (があるの) を知らなかったから	7
馴染みがあるから	2
友達にやってもらった	1
電話番号を変更したから	1
一度試したら早かったから	1
特に覚えていない	5

第 5 章

国の異なる全ての人を対象にした調査

5.1 概要

本実験は、様々な国に住む性別や年齢の異なる 23 名のユーザを対象に行った。ユーザの基本情報を表 5.1 に示す。アンケート項目は基本情報について、SeBIS について、スマートフォンを使用した多要素認証についての 3 種類からなる。それぞれの質問の回答データを収集し、国の違いとセキュリティ意識や多要素認証方法の選択の関連について考察する。

5.2 国の違いとセキュリティ意識の違い

本実験に協力したユーザの国について、アメリカ州、アジア州、ヨーロッパ州という区分に分けた際のセキュリティ意識の違いを考察する。アメリカ州には USA とブラジル国籍のユーザが合計 6 名、アジア州には台湾と韓国と日本のユーザが合計 10 名、ヨーロッパ州にはイタリアとフランスとスイスのユーザが合計 6 名属している。それぞれの州ごとの SeBIS の得点分布を表 5.1 の箱ひげ図に示すが、このグラフから、州によってセキュリティ意識が大きく異なることはないことがわかる。そこで、日本とその他の国との間で比べるため、4 章の実験で得た 57 名の日本人のデータと本実験の日本人 3 人のデータと、その他の国の 20 人のデータを用いてセキュリティ意識の違いを調べた。その結果を表 5.2 に箱ひげ図で示す。このグラフから、日本の外れ値を除くと海外のデータの方がやや高い値を示すことが分かる。海外のもの最高値が 70 で最低値が 37、平均値が 53.6 なのに対し、日本のものは最高値が 64、最低値が 34、平均値が 48.7 であった。しかし、データの数の違いや、日本のデータは学生が大半を占めるのに対し、海外のデータは社会人を含んでいることから、海外と日本の違いであるとは言い切れない。

5.3 認証方法の人気度・セキュリティ意識との関連

最も使用する頻度の高い認証方法の統計を表 5.3 に示す。最も多く使用する認証方法を一つ回答せよという質問の結果、SMS コードによる認証が 11 名で全体の 47.8% と最も多く、次に生体認証が 9 名で全体の 39.1%、次にモバイルアプリによる認証が 3 名で全体の 13% であった。やはり海外でも SMS によるワンタイムパスワードを利用した認証が主流であることがわかった。次に多かった生体認証については、SMS やモバイルアプリでの認証の中でショートカットとして使用している可能性が高いため、あまりあてにならない結果だと推測する。モバイルアプリの認証を選択した人数は 3 名であったが、そのうち日本人が 1 名、台湾人が 2 名で全員

表 5.1 国別ユーザ実験基本情報

		人数 (名)	割合 (%)
性別	男性	6	26.1
	女性	17	73.9
年齢	～19 歳	3	13.0
	20～24	10	43.5
	25～30	3	13.0
	31～	7	30.4
国	USA	5	21.7
	台湾	7	30.4
	イタリア	2	8.7
	韓国	1	4.3
	フランス	2	8.7
	スイス	2	8.7
	ブラジル	1	4.3
	日本	3	13
社会人経験の有無	ある	17	73.9
	ない	6	26.1

20～24 才であった。このことから、日本や台湾などのアジア圏の方がアプリの浸透があるとも考えられる。また、図 5.4 にはモバイルアプリによるワンタイムパスワードと SMS コードによるものごとの SeBIS の得点分布を箱ひげ図で示す。生体認証については前述のとおりあてにならないと判断したため比較対象から外した。このグラフから、それぞれを選んだ人の SeBIS の得点分布はさほど変わらないことが分かる。モバイルアプリを選んだ人の平均は 50.6 であり、最低値は 46, 最高値は 58, SMS コードを選んだ人の平均値は 52.9 であり、最低値は 42, 最高値は 63 である。しかしモバイルアプリのデータの数が 3 であるのに対し SMS コードのものは 11 でデータの数に差があることから、あまり信頼性の高い結果ではないと考える。

5.4 考察

本実験結果から、国の違いによるセキュリティ意識の違いや、多要素認証方法の違いは見られないことが分かった。どの国でもモバイルアプリよりも SMS コードによるワンタイムパスワードを第二要素とした認証がまだまだ主流であると考えられる。

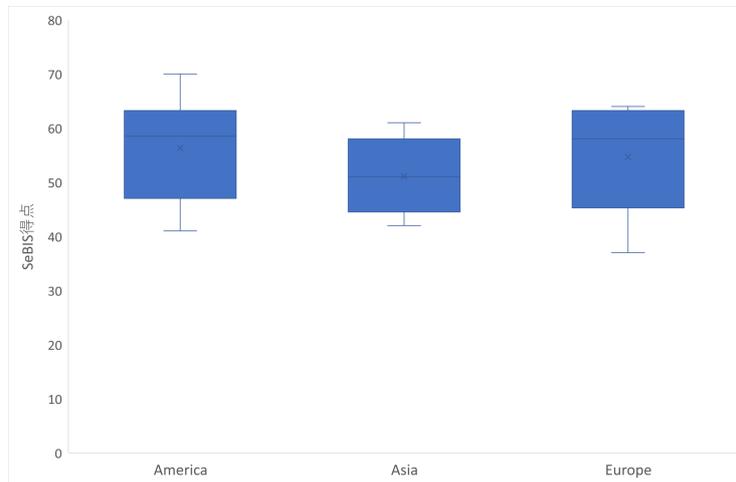


図 5.1 州別の SeBIS 得点分布箱ひげ図

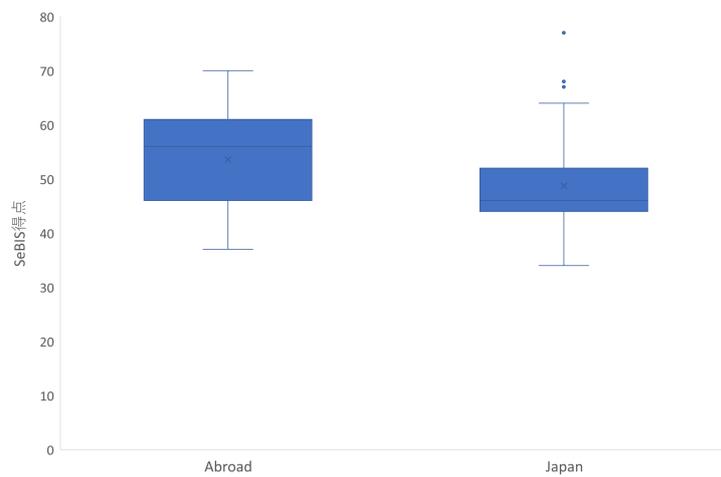


図 5.2 日本とその他の国の SeBIS 得点分布箱ひげ図

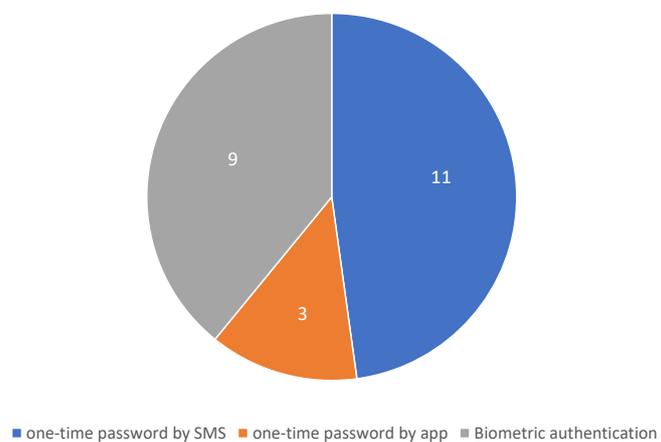


図 5.3 人気な認証方法

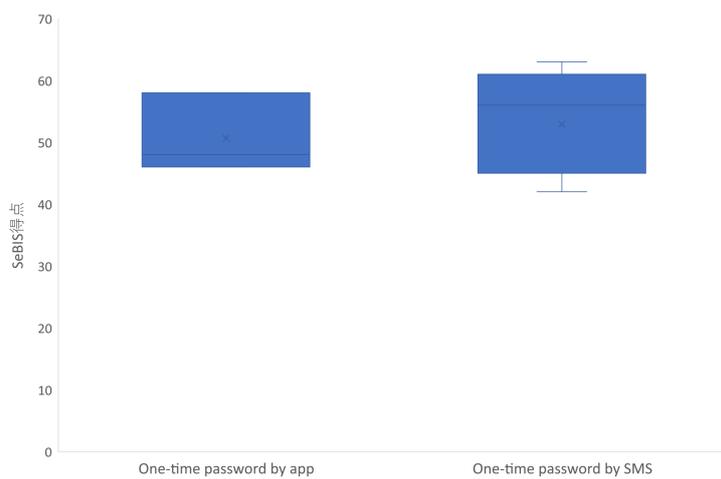


図 5.4 認証方法別の SeBIS 得点分布箱ひげ図 (国別)

第6章

おわりに

本実験から、学部や国などの人の属性と選択する認証方式には相関が見られないこと、主流の認証方式がどの国でも SMS コードであることが分かった。また、モバイルアプリを認証方法に選ぶ人のセキュリティ意識は SMS コードのものよりも高い傾向が見られた。しかし中間者攻撃の認知や安全面が必ずしもモバイルアプリを選択する理由ではなく、「他の方法を知らなかったから」等の消極的な理由が主であった。また、SMS コードを認証方法に選択する理由は「慣れ親しんでいるから」がほとんどであった。これらのことから、モバイルアプリでの認証方法の安全性や利便性を広く認知させること、モバイルアプリでの認証を推奨することが必要であると結論付ける。

2024.02 Aoba Koike

参考文献

- [1] 明治大学, MeijiMail 多要素認証 (<https://www.meiji.ac.jp/isc/mm-mfa/mfa.html>, 2023.12 参照)
- [2] 野口, 大瀧, 山本, 西原, 外岡, “レベル分けによる多要素認証の要否の実現と認証連携の拡張”, 学術情報処理研究 Mo.24 JACN 2020
- [3] Lyastani, Backes, Bugiel, “A Systematic Study of the Consistency of Two-Factor Authentication User Journeys on Top-Ranked Websites”, Network and Distributed System Security (NDSS) Symposium 2023
- [4] Nethanel Gelernter, Senia Kalma, Bar Magnezi, Hen Porcilan, “The Password Reset MitM Attack”, IEEE Security and Privacy 2017,
- [5] 柴山, 菊池, “多要素認証を悪用したパスワードリセット手法 PRMitM 攻撃の被害を増加させる新たな要因の調査”, 情報処理学会創立 60 周年記念 第 82 回大会, 2020
- [6] 笹, 菊池 “二要素認証を悪用したパスワードリセット手法 PRMitM の影響評価”, 2018 Symposium on Cryptography and Information Security Niigata, Japan, Jan. 23 - 26, 2018
- [7] Egelman, Peer, “Scaling the Security Wall, Developint a Security Behavior Intentions Scale(SeBIS)”, CHI2015, pp2873-2882

謝辞

本研究を行うにあたり，多くの方より御指導いただきました．特に，多大なる御指導を受け賜りました，明治大学総合数理学部先端メディアサイエンス学科，菊池浩明教授に深く感謝申し上げます．予備実験，本実験に協力してくださった明治大学学生の皆様並びに様々な国に住むすべての方々に深く感謝の意を表するとともに，謝辞とさせていただきます．

付録 A

k 匿名化ツールである Nithin Prabhu の改良と評価

A.1 はじめに

A.1.1 背景

近年、膨大な個人データを扱うシステムの運用上、個人を特定できないようにデータを匿名化すること多い。しかし、複数のデータの照合で個人を特定できてしまうという問題が度々指摘されている。[6]

匿名加工方法の 1 つとして、 k 匿名化というものが挙げられる。これは、同一属性のデータを、 k 件以上になるようデータを変換し、個人が特定される確率を k 分の 1 以下に低減して特定を困難にするという方法だ。[6]

これらの背景を踏まえ、本研究ではその k 匿名化について、代表として 1 つのツールについて調査することで実態を検証する。本研究では、 k 匿名化ツールである Nithin Prabhu について、プログラムの改良と有用性の調査を行う。Nithin は課題として、匿名加工後に出力される任意の 2 つの準識別子についての分布グラフの、度数が一目で分からないというものがあった。

A.1.2 目的

本研究では、 k 匿名化ツールである Nithin Prabhu のプログラム改良と評価を目的とする。プログラム改良については、1 章で述べた課題を踏まえ、度数をグラフ上で把握できるように可視化プログラムを追加する。また評価については、最も有用性の高い k を調査することを目的として、 k の値を変化させたときの準識別子の平均値を元データとの差分で評価する。

A.1.3 使用ツール

本ツールは k 匿名化の手法として、Mondrian algorithm を使用しており、任意の 2 つの準識別子についてのグループ分割境界のグラフを出力する。Mondrian algorithm は、個別データを、 k 個以上のレコードを持つグループ単位で属性値を一般化することで k 匿名化を行う。このツールのシステム構成図を図 A.1 に示す。また、システム改良前の出力グラフを図 A.2 に示す。クラスタの範囲を示すレクトとレコードがある場所を示す点があり、度数に比例した濃さの色がレクトに反映されている。

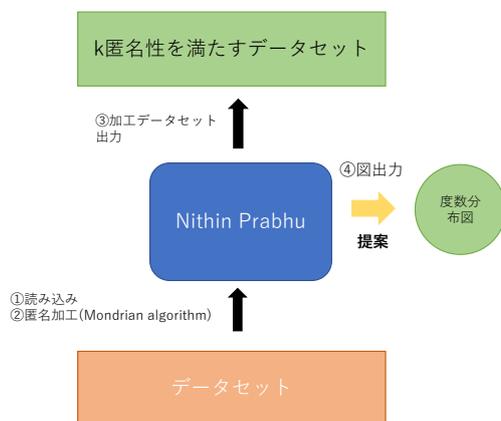


図 A.1 システム構成図

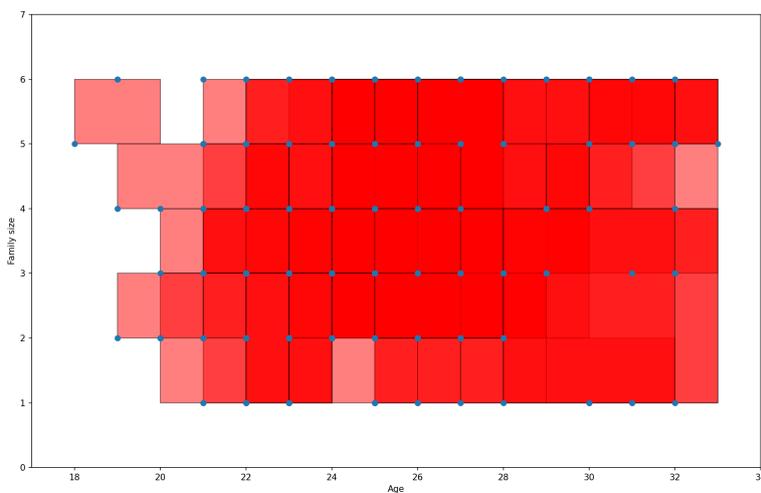


図 A.2 実行結果

A.1.4 実行データ

本研究では匿名加工を行う際に, onlinedeliverydata を使用する. このデータはオンラインデリバリーを使用した人の個人情報を含む. ただし, 本研究ではその中でも必要なカラムである計 6 カラムに削ったデータを使用する. 使用データのカラムの一覧を表 A.1 に表す. また準識別子は Age, Gender, Marital Status, Occupation, Family size とし, Monthly Income をセンシティブ属性とする. ここで, 準識別子を間接的に個人を識別できる属性, センシティブ属性を他人にむやみに知られたくない属性と定義する. グラフのプロットには, 準識別子

表 A.1 Onlinedelivery データの一部

添え字	Age	Gender	Marital Status	Occupation	Family size	Monthly Income
0	20	Female	Single	Student	4	No Income
1	24	Female	Single	Student	3	Below Rs.10000
2	22	Male	Single	Student	3	Below Rs.10000
...
387	23	Male	Single	Student	5	No Income

に連続値である Age と Family size を採用する.

A.2 先行研究

A.2.1 先行研究 1

1 つ目は, 2016 年に小栗, 黒政, 松井らによって発表された“匿名加工・再識別コンテストを通じた情報流通プラットフォームの検討” [1] である. 概要は, 匿名加工情報の登録と評価を行うためのプラットフォームの構造と評価方法に求められる要件についての検討している.

A.2.2 先行研究 2

2 つ目は, 2016 年に南, 千田らによって発表された“再識別リスク評価と匿名化の展望” [2] である. 概要は, 匿名化データの再識別リスクについて, 現状の指標との対応付けや課題, そして対策について考察している.

A.3 提案システム

A.3.1 概要

本研究では, Nithin Prabhu の k 匿名化とクラスタの可視化プログラムを基にして, 各 k 匿名性を満たすクラスタごとのレコード数をグラフに出力する. k の値はデフォルトで設定されていた 3 とする. グラフは Age と Family Size の散布図である.

A.3.2 提案方法

k 匿名性を満たすクラスタを作成するたびに, そのクラスタ内のレコード数を記録し, プロットさせる関数に渡す. クラスタを多角形レクトとしてプロットする際, そのレクトの中央に度数を描画する. この数値を A とし, レクトの右上の座標を (x_r, y_r) , 左下の座標を (x_l, y_l) とすると, A を描画させる位置の座標は $((x_r + x_l)/2, (y_r + y_l)/2)$ と表せる. A は, アルゴリズムに基づいてまず, 属性 1(図 A.3(1) の x) 中央値で分割する. 次に属性 2(図 A.3(2) の y) を分割する. これを繰り返す.

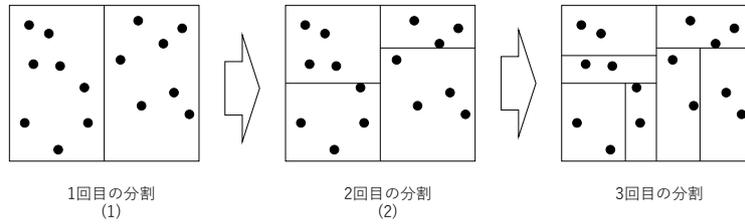


図 A.3 Mondrian アルゴリズムの動作イメージ

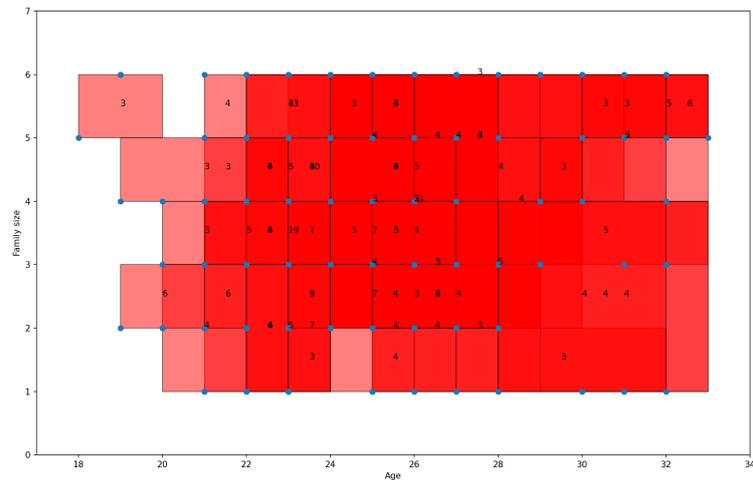


図 A.4 本プログラムの実行結果

A.3.3 実行結果

本プログラムの出力を図 A.4 に示す。本プログラムでは、レクトの中央に対応するグループごとの度数を表示している。例えば、Family size = [5-6], Age = [18-20] では度数が 3 となる。

A.3.4 有用性評価

本研究では、 k の値を変化させたときの準識別子の平均値と元データの平均値との差で有用性を定義する。ある準識別子 a について、加工前のデータを $S = \{s_1, s_2, s_3, \dots, s_n\}$, 加工後のデータを $T = \{t_1, t_2, t_3, \dots, t_N\}$

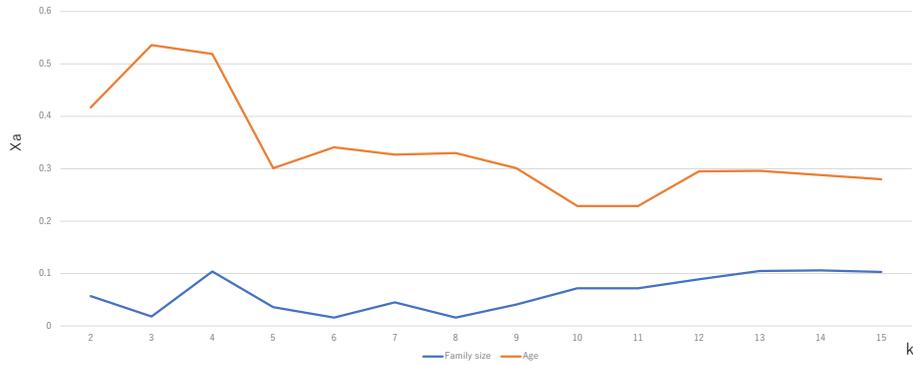


図 A.5 k についての有用性の分布

とすると, 求める差分 X_a は

$$X_a = \left| \frac{1}{n} \sum_{i=1}^n s_i - \frac{1}{N} \sum_{j=1}^N t_j \right|$$

で定義される. X_a を全ての a について平均したものを有用性と定める. 本研究では, k の値を 2 から 15 まで変化させ, 最も X_a が小さい時の k を有用性が最も高い時のものとする.

A.3.5 評価

k の値と差分の関係を図 A.5 に示す. Age については k が 10, 11 のとき, Family size については k が 6, 8 の時に最も有用性が高くなるという結果となった.

A.3.6 考察

本改変プログラムには, 度数を可視化することはできたが, グラフとしては見づらいものとなった. 有用性の評価については, 図 A.5 より Family size と k の値には僅かに正の相関が見られた. Age と k の値にはそれよりも僅かに負の相関が見られた. また, 準識別子ごとに有用性が高くなる時の k の値は異なっていた.

A.4 おわりに

改変プログラムについて, 領域をプロットすると同時に度数を表示するプログラムだと字が重なってしまう部分があり, 改善が必要である. 有用性の評価について, 評価方法が単純で問題もあり, 今後拡張を検討する.
2023.02 Aoba Koike.

参考文献

- [1] 小栗, 黒政, 松井, “匿名加工・再識別コンテストを通じた情報流通プラットフォームの検討”, 情報処理学会研究報告, 2016 Information Processing Society of Japan, pp.1-8, 2016.
- [2] 南, 千田, “再識別リスク評価と匿名化の展望”, Computer Security Symposium 2016, pp.166-172, 2016.
- [3] 伊藤充司, “インシデント損害金額を推定するウェブサイトの開発と評価”, 明治大学総合数理学部卒業論文, 2021.
- [4] NithinPrabhu/K-Anonymity (<https://github.com/Nuclearstar/K-Anonymity>, 2022年5月参照)
- [5] Kaggle/Online Food Delivery Preferences-Bangalore region(<https://www.kaggle.com/datasets/benroshan/online-food-delivery-preferencesbangalore-region>, 2022年8月参照)
- [6] 伊藤秀樹 “データ活用の利便性と安全性を両立させる 匿名化技術の最新動向とその課題”, 2014 NII Today No.64, pp10-11
- [7] 千田, “個人特定のリスクを低減させる匿名化技術”, オペレーションズ・リサーチ, 2016.
- [8] 古川, 佐古, “パーソナルデータ秘匿化技術のサービス活用”, Society for Serviceology, 2020年6月