既知ユーザ攻撃によるユーザ情報の漏洩リスクを低減した条件マッチング連係方式

竹之内隆夫† 南澤岳明† 伊東直子†

近年，様々なサービス事業者で膨大なユーザ情報が収集されている．今後は，これらユーザ情報を組み合わせたサービスが生まれてくると期待されている．著者らは，比較的計算量が少ない方式で，多数の事業者がもつ膨大なユーザ情報を開示せずに，条件に合致するユーザを抽出する方法について研究しており，条件マッチング連係方式を提案した．しかし，この方式には，ある事業者が一部のユーザについてのユーザ情報を背景知識として持っていると，他の事業者のユーザ情報を推測出来てしまうという推測攻撃が存在する．著者らは，これを既知ユーザ攻撃と呼んでいる．そこで，本論文では，条件マッチング連係方式を拡張し，既知ユーザ攻撃によるユーザ情報の漏洩リスクを低減する新たな連係方式を提案する．また，漏洩リスクを算出するための評価式を用いて本方式の評価を行い，本方式によって既知ユーザ攻撃によるユーザ情報の漏洩リスクが低減することを示す．

Combined Rule Matching with Reducing the Risk of User Information Leakage by Known-User Attack

TAKAO TAKENOUCHI† TAKEAKI MINAMIZAWA† NAOKO ITO†

Recently, many service providers collect vast amounts of user information. It is expected that the user information stored in different service provider is combined and used together, in order to provider new services. The authors have been researching the method to use the combined user information without disclosing user privacy. And, the authors proposed Combined Rule Matching method which reduces the user information leakage risk and the high calculation cost. In this method, however, there is a risk that a participating party may be able to inference the user information of all the selected users if the participating party knows some of the user information which is owned by other participating parties. We call this attack as Known-User Attack. This paper introduces an extended method of the Combined Rule Matching which reduces the risk of user information leakage by the Known-User Attack. Also, it introduces a new measure of the leakage risk and provides the evaluation result that shows that the proposed method can reduce the risk of user information leakage.

†日本電気株式会社 サービスプラットフォーム研究所
NEC Corporation, Service Platforms Research Laboratories.

1. はじめに

近年，インターネット上やクラウド上のサービス事業者では，ユーザに適したサービスを提供するために，多くのユーザ情報を収集している．今後は，これらのユーザ情報は，単一の事業者内で利用されるだけにとどまらず，様々なサービス事業者に存在するユーザ情報が組み合わせて利用され，新たなサービスが生まれてくると期待されている[1].

これらのユーザ情報はプライバシ情報であるため，プライバシを保護して利用する必要がある．例えば，アクセスコントロールのような開示制御による方法では，必要最小限のプライバシ情報の開示で済ますことができるが，それでも一度開示された情報は，開示先の事業者内部での不正や誤操作等の恐れがあり，情報漏洩のリスクが増すことになってしまう．また，Multi-Party Computation[2]のような暗号技術を利用した方法では，情報を一切開示せずに任意の計算が可能であるが，膨大なユーザ情報に対して処理するには計算量が多くなってしまう[3].

そこで，著者らは，比較的計算量が少ない方式で，多数の事業者がもつ膨大なユーザ情報を他の事業者へ開示せずに，指定されたユーザ情報の条件に合致するユーザを抽出し，メッセージを配信する方法について研究しており，条件マッチング連係方式を提案した[4][5]．この既提案方式を用いれば，例えば，通信販売サイトが持つ購買履歴による嗜好情報と，地図提供サイトが持つ位置情報を組み合わせた「『ゲームが好き』で『渋谷にいる』人」という条件に合致するユーザに対して，ターゲティング広告を行うことが出来る．

しかし，既提案方式では，ある事業者が，一部のユーザについてのユーザ情報を背景知識として持っていると，その事業が，他の事業者のユーザ情報を推測出来てしまう．著者らは，この推測方法による攻撃を，既知ユーザ攻撃と呼んでいる．本論文では，条件マッチング連係方式を拡張し，既知ユーザ攻撃による情報推測のリスクも低減させた方式を提案し，評価を行う．

本論文は，以下のような構成になっている．まず 2 章で，著者らが既に提案している条件マッチング連係方式について説明する．次に，3 章で，既知ユーザ攻撃と既知ユーザ攻撃によってどのようにユーザ情報が推測されるかについて説明する．そして，4 章で，既知ユーザ攻撃によるユーザ情報の推測を軽減するための方式として，ユーザ ID の送受信を仲介する ID ルータ機能を用いた改良版条件マッチング連係方式を提案する．5 章では，提案した改良版条件マッチング連係方式によって，どれだけユー

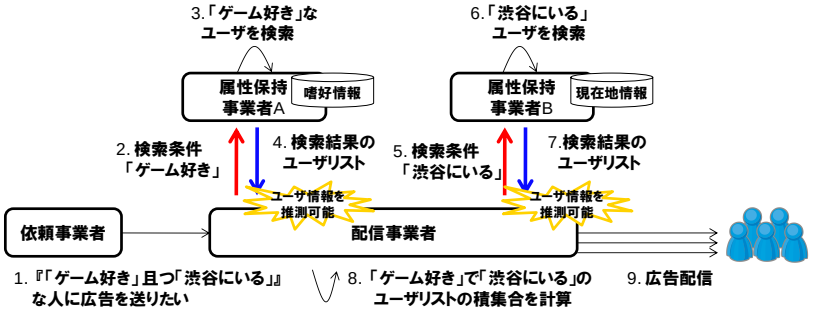


図 1 従来の連係方式(検索条件と検索結果からユーザ情報の推測が可能)

ザ情報の推測リスクが軽減したかを評価する．最後に 6 章で，本論文をまとめる．

2. 条件マッチング連係方式

2.1 ユーザ情報の連係とプライバシ

従来，複数の事業者が連係して，各事業者が持つユーザ情報を利用してユーザの検索を行う場合には，ある事業者が，各事業者が持つユーザ情報の複製を取得したり，検索条件を与えて，その検索結果を取得したりしていた．例えば，図 1 に示したような，ある条件に合致するユーザに広告を送るような場合，広告配信を行う配信事業者が，ユーザ情報を保持する属性保持事業者に対して検索条件を送信し，そして，検索結果のユーザのリストを取得していた．この方式では，配信事業者は，検索条件と検索結果から，容易にユーザ情報を知ることができる．例えば，「ゲーム好き」の人を検索してほしいという検索条件に対して，user1,user2 という検索結果が返ってきた場合，user1 と user2 はゲームが好きということを知れてしまう．

2.2 マッチング条件と条件マッチング結果の分離による連係方式

そこで，著者らは，検索条件と検索結果を分離した連係方式として，条件マッチング方式を提案した[4][5]．この連係方式を用いることで，従来の連係方式のような，検索条件と検索結果からの容易なユーザ情報の推測を困難にでき，ユーザ情報の漏洩リスクを軽減した連係が実現できる．

この連係方式は，「条件設定機能」と「条件マッチング機能」により構成される(図 2)．「条件設定機能」は，配信依頼事業者に配置され，「条件マッチング機能」はユーザ情報を保持する各事業者内に配置される．まず，依頼事業者が，「条件設定機能」に，どのようなユーザに配信したいかの条件を示したマッチング条件を設定する．「条件設定機能」は，そのマッチング条件を各事業者で処理する個別のマッチング条件に分離

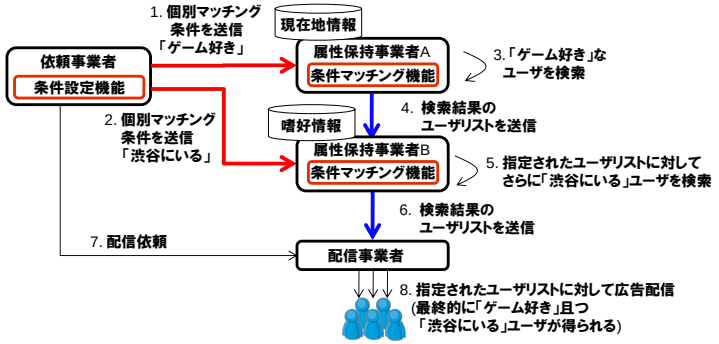


図 2 条件マッチング連係方式

する．次に，分離した個別のマッチング条件を，各属性保持事業者の「条件マッチング機能」へ送信する．「条件マッチング機能」は，受信した個別のマッチング条件に合致するユーザを抜き出すというマッチング処理を行い，結果のユーザのリストを次の属性保持事業者へ送る．次の属性保持事業者の「条件マッチング機能」では，指定されたユーザのリストを母集団として，同様にマッチング処理を行う．そして，結果を次の属性保持事業者へ送信する．このように，各「条件マッチング機能」で徐々にユーザをマッチング処理していくことで，最終的に依頼事業者が指定したマッチング条件に合致したユーザのグループが作成され，広告が配信される．

この連係方式では，「条件マッチング機能」へマッチング条件を設定する依頼事業者と，「条件マッチング機能」からマッチング結果を受け取る事業者が異なるように連係している．これは，条件を与えて，その条件に合致するユーザリストを取得してしまうと，ユーザ情報が他の事業へ知られてしまうという問題を回避するためである．このように，条件マッチング連係方式によって，ユーザ情報の漏洩リスクを軽減した連係が実現できる．

2.3 連係する事業者間でのユーザリストの送受信のためのユーザ ID 変換

条件マッチング連係では，条件に合致したユーザのユーザ ID のリストを，他の属性保持事業者や配信事業者へ送信する．一般的に，異なる事業者では異なるユーザ ID が利用されているため，ユーザ ID のリストを送信するためには，送信先の事業者のユーザ ID に変更する必要がある．

送信先の事業者のユーザ ID へ変更するために，条件マッチング連係では，SAML Identity Provider[6]や OpenID Provider[7]のような ID 管理機能を利用する．ID 管理機能は，ID 管理事業者が運営しており，各属性保持事業者のユーザ ID と，ID 管理事業者のユーザ ID の紐付け情報(ID 連携情報)を保持している．属性保持事業者は，この ID

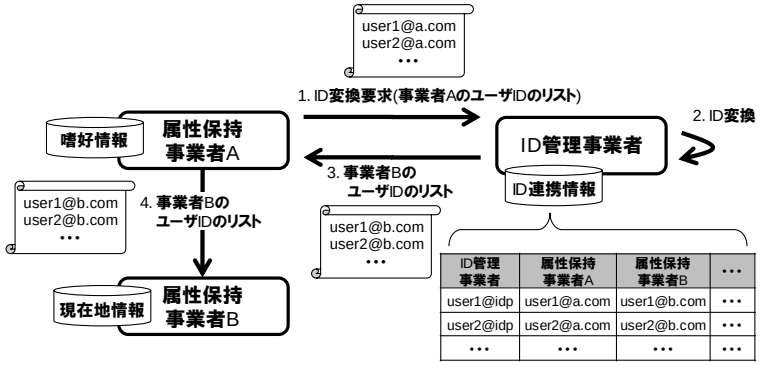


図 3 事業者間でのユーザリストの送受信のためのユーザ ID 変換

管理機能を用いて，送信先の事業者のユーザ ID に変換し，変換後のユーザ ID を送信先の事業者へ送信する(図 3).

3. 既知ユーザ攻撃

3.1 既知ユーザ攻撃によるユーザ情報の推測の概要

既に提案している条件マッチング連係方式は，各事業者が保持するユーザ情報が，別の事業者に渡ることを防ぐことで，プライバシ情報の漏洩を防ぐものである．しかし，実際の事業者連携では，ある程度の攻撃を考慮した設計を行う必要がある．本論文では，各事業者は，他の事業者のユーザ情報の推測が可能であれば，積極的に推測攻撃を行うという前提とする．ただし，他の事業者と結託をしたり，プロトコルを逸脱したりしてまで攻撃を行うことは無いという前提とする．

このように各事業者が攻撃を行うという前提の場合，2 章で説明した条件マッチング連係方式には，もし，ユーザリストを受信する事業者が，受信したユーザリストに背景知識のあるユーザが含まれていると，他の事業者のユーザ情報を推測出来てしまうという推測方法が存在する．これは，受信したユーザリストに含まれていた，背景知識のあるユーザから，マッチング条件を推測でき，結果として背景知識のないユーザのユーザ情報も推測できてしまうからである．著者らは，この背景知識のあるユーザを「既知ユーザ」と呼び，この推測方法による攻撃を「既知ユーザ攻撃」と呼ぶ．

既知ユーザ攻撃の詳細を，例を用いて説明する．例えば，ユーザリストを送信した事業者は「好み」についてのユーザ情報を保持している事が公開されており，さらに，ユーザリストを受信した事業者が，あるユーザの「好み」が「ゲーム」である事を背景知識として知っていたとする．すると，ユーザリストを受信した事業者は，受け取ったユーザのリストにその既知ユーザが含まれていた場合，既知ユーザの「好み」が「ゲーム」であることから，受け取ったユーザのリストは「好み＝ゲーム」という条件に合致するユーザのリストであると推測が出来てしまう．つまり，ユーザリストを受信した事業者は，受け取ったユーザリストの全ユーザの「好み」が「ゲーム」であると推測できる．よって，ユーザリストを受信した事業者に対して，ユーザリストを送信した事業者のユーザ情報が漏洩していることになってしまう．

既知ユーザ攻撃は，ユーザリストを受け取る事業者において可能である．つまり，属性保持事業者だけでなく，配信のためのユーザリストを受け取る配信事業者や，ID 変換のためにユーザリストを受け取る ID 管理事業者も，既知ユーザ攻撃が可能である．

3.2 既知ユーザ攻撃によるユーザ情報の推測

既知ユーザ攻撃により，以下の 2 つのユーザ情報の推測が行われる．これらの推測の内容について，3.2.1 節と 3.2.2 節で説明する．

(a) 単一既知ユーザ推測
(b) 合致既知ユーザ推測

また，本論文では，既知ユーザ攻撃におけるユーザ情報の推測方法を検討するのを容易にするため，条件マッチング連係方式では以下のような前提を置く．

- 各属性保持事業者には，同一のユーザが存在する．
- 各属性保持事業者は，全ユーザのユーザ情報を 1 種類だけ持つ．
- 各属性保持事業者がどのようなユーザ情報の種類を持つかは，公開されている．
- 依頼事業者が設定できる条件は，ユーザ情報の種類毎に 1 つの値とする．例えば，「年齢= 20 代 or 30 代」という指定はできない．
- 既知ユーザ攻撃を行う事業者は，直前の上位の属性保持事業者が持つユーザ情報の推測を行うとする．

3.2.1 単一既知ユーザ推測

既知ユーザ攻撃は，受け取ったユーザリストに 1 名でも既知ユーザが含まれていると，既知ユーザのユーザ情報を利用して，受け取ったユーザリストに含まれる他のユーザのユーザ情報を，有る程度の確信度で推測する事ができる．図 4(a)に，受け取ったユーザリストに既知ユーザが 1 名でもいた場合における，ユーザ情報の値の推測の例を示す. この例では, ユーザリストを受け取る事業者(属性保持事業者, 配信事業者, ID 管理事業者)が, 既知ユーザとして user1 と user2 のユーザ情報を知っていたとする．この時，受け取ったユーザリストに user1 が含まれていた場合，少なくとも受け取ったユーザリストは，直前の上位の属性保持事業者によって，既知ユーザの user1 の各ユーザ情報に関するいずれかのマッチング条件で，合致されたユーザのリストであると推測できる．この例では，直前の上位の属性保持事業者は，「年収」か「年齢」か「好み」か「現在地」のいずれかをもつ属性保持事業者である．user1 の「年収」が「400

万円台」であるので，仮に，直前の上位の属性保持事業者が「年収」を保持する事業者であった場合は，受け取ったユーザリストの user5 と user6 は，「年収」が「400 万円台」であることになる．「年齢」や「好み」についても同様な推測が可能である．ただし，この例では user1 の「現在地」については不明であるため，「現在地」についての推測は出来ない．

このように，「単一既知ユーザ推測」によって，有る程度の確信度での直前の上位の属性保持事業者のユーザ情報の推測が可能である．

3.2.2 合致既知ユーザ推測

既知ユーザ攻撃には，受け取ったユーザリストに既知ユーザが複数いた場合に，「単一既知ユーザ推測」よりもさらに高い確信度で，直前の上位の属性保持事業者のユーザ情報の推測が可能である．図 4(b)に，「合致既知ユーザ推測」の例を示す．この図に示したように，例えば，ユーザリストを受け取る事業者が，既知ユーザとして user1 と user2 と user3 を知っていたとする．この時，もし受け取ったユーザリストに，user1 と user3 が含まれていた場合，user1 と user3 は，直前の上位の属性保持事業者において，なんらかの条件マッチングが行われているので，そのマッチング条件に関するユーザ情報が一致しているはずである．この例では，「年収」と「年齢」が一致しており，「好み」が一致していない．つまり，上位の属性保持事業者において，「好み」に関する条件マッチングは行われていない事がわかる．これにより，直前の上位の属性保持事業者では，「年収」か「年齢」か「現在地」のいずれかを持つ属性保持事業者であることが解る．

先ほどの例の「単一既知ユーザ推測」では，直前の上位の属性保持事業者の候補は，「年収」か「年齢」か「好み」か「現在地」を持つ 4 つの属性保持事業者のいずれかで有ったが，この例の「合致既知ユーザ推測」では，「年収」か「年齢」か「現在地」を持つ 3 つの属性保持事業者のいずれかである．よって，直前の上位の属性保持事業者の候補が少ない分，直前の上位の属性保持事業者の推測が容易になる．

このように，「合致既知ユーザ推測」では，「単一既知ユーザ推測」よりもさらに高い確信度で，直前の上位の属性保持事業者のユーザ情報の推測が可能である．

3.2.3 直前の上位事業者の確定による確信度 100%の推測

「単一既知ユーザ推測」や「合致既知ユーザ推測」推測では，有る程度の確信度で，直前の上位の属性保持事業者におけるユーザ情報の推測が可能であったが，さらに，ユーザリストの送信元の事業者が解る事で，確信度 100%での推測が可能になってしまう．既存の条件マッチング連係では，属性保持事業者から，直接次の属性保持事業者や配信事業者へユーザリストを送信する．そのため，ユーザリストを受け取った事業者は，直前の上位の属性保持事業者が解る．この事から，例えば，先ほどの「単一既知ユーザ推測」の例の場合，直前の属性事業者が「年収」を保持する事業者であった場合は，ユーザリストを受け取った事業者は，user5 と user6 の「年収」が「400 万円台」であることを，確信度 100%で推測出来てしまう．

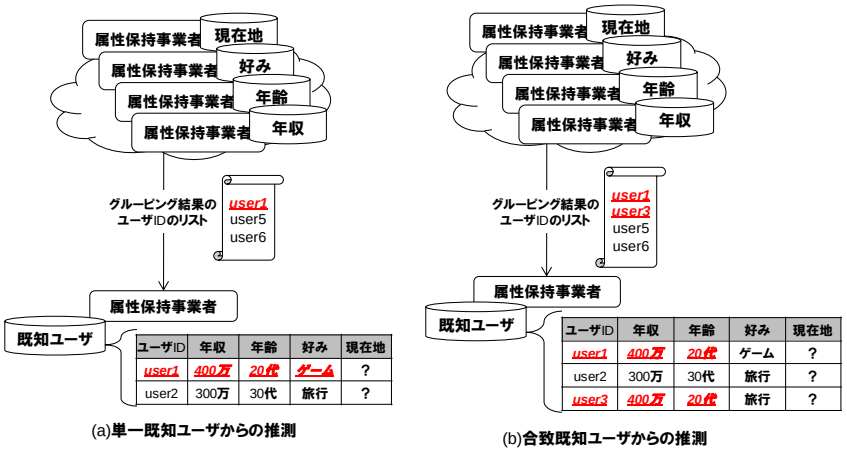


図 4 既知ユーザ攻撃によるユーザ情報の推測の例

4. ID ルータ機能による条件マッチング連係方式の提案

4.1 既知ユーザ攻撃の軽減の方針と ID ルータ機能

3.2.3 節で説明したように，「単一既知ユーザ推測」や「合致既知ユーザ推測」において，直前の上位の属性保持事業者が解ると，ユーザ情報を 100%の確信度で推測出来てしまっていた．つまり，直前の上位の属性保持事業者を解らなくすることで，「単一既知ユーザ推測」や「合致既知ユーザ推測」によるユーザ情報の推測を困難にすることができる．

そこで，ユーザ情報の推測の確信度を減らすために，ユーザリストを属性保持事業者から，次の属性保持事業者や配信事業者へ送信する際に，ID 管理事業者が仲介する方式を提案する．この方式では，ID 管理事業者は単にユーザ ID の変換を行うだけでは無く，変換したユーザ ID を，指定された事業者へ送信する．著者らは，そのような機能を「ID ルータ機能」と呼んでいる．ID ルータ機能を用いて，条件マッチング連係方式を拡張する事で，ユーザリストを受信した事業者が，直前の上位の属性保持事業者を解らなくすることができるので，「単一既知ユーザ推測」や「合致既知ユーザ推測」によるユーザ情報の推測の確信度を減らすことができる．

4.2 ID ルータ機能による改良版条件マッチング連係方式の詳細

図 5 に，ID ルータ機能の動作を示す．この図に示したように，ID ルータ機能を用いた改良版条件マッチング連係方式では，属性保持事業者 A がユーザリストを送信する際に，属性保持事業者 A が，直接，属性保持事業者 B に送信するのではなく，ID

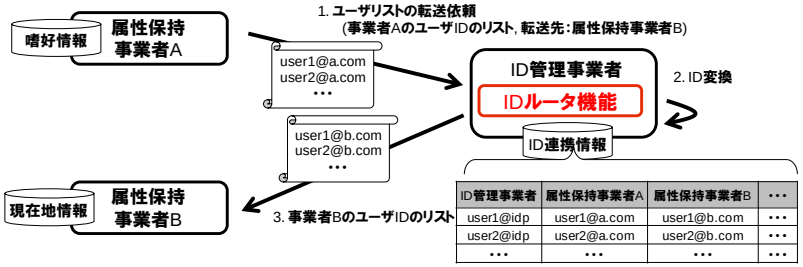


図 5 ID ルータ機能によるユーザ ID 変換とユーザリスト送信の例

管理事業者の ID ルータ機能に，ユーザリストの送受信を仲介させる．ID ルータ機能は，送信元のユーザ ID である属性保持事業者 A のユーザ ID のリストを受け取り，そのユーザ ID を送信先の事業者である属性保持事業者 B のユーザ ID に変換する．そして，変換したユーザ ID のリストを，送信先の属性保持事業者 B へ送信する．このように，ID ルータ機能が仲介することで，属性保持事業者 B に，ユーザリストが属性保持事業者 A から送られてきたことを知られずに，ユーザリストを渡すことができる．

ID ルータ機能は，ID 管理事業者に配置されるので，新たな既知ユーザ攻撃を行う事業者が増える事はない．ID 管理事業者は，2.3 節で説明したように，既提案の条件マッチング連係方式でも，ID 管理事業者では ID 変換のためのユーザリストの送受信が必要であった．つまり，ID 管理事業者に ID ルータ機能を導入することによって，新たに既知ユーザ攻撃が行える事業者が増えるようなことは無い．

図 6 に, ID ルータ機能を用いた改良版条件マッチング連係方式の全体の構成を示す．この図に示したように，ID ルータ機能は，属性保持事業者から次の属性保持事業者へユーザリストを送信する際や，属性保持事業者から配信事業者へユーザリストを送信する際に，ユーザリストの送受信を仲介する．このように，ID ルータ機能が仲介することで，ユーザリストを受け取る属性保持事業者や配信事業者は，直前の上位の属性保持事業者を知ることを防ぐことが出来る．よって，属性保持事業者や配信事業者における，「単一既知ユーザ推測」や「合致既知ユーザ推測」による，直前の上位の属性保持事業者のユーザ情報を 100%の確信度で推測されることを防ぐことができる．

5. 評価

本章では，ID ルータ機能を用いた改良版条件マッチング連係機能の評価を行う．まず，5.1 節で，評価の指標となるリスク値の計算式について説明する．次に，5.2 節で，評価に用いたデータセットについて説明する．そして，5.3 節で，単一の属性保持事業者に対する既知ユーザ攻撃の推測リスクについて評価する．続いて，5.4 節で，システム全体の既知ユーザ攻撃の推測リスクについて評価する．

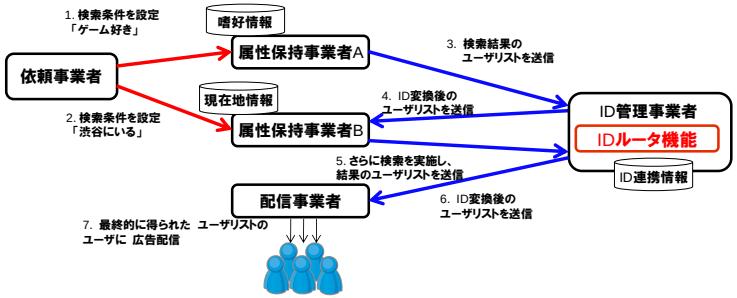


図 6 ID ルータ機能による条件マッチング連係方式の全体構成

5.1 ユーザ情報の推測リスクの計算式

本節では，既知ユーザ攻撃によるユーザ情報の推測リスクの式を定義する．ISMS(Information Security Management System)[8]のリスク値の考え方を参考にすると，リスク値は『「被害の大きさ」×「被害が起こる可能性」』で表される．そこで，本論文では,「被害の大きさ」を, 推測が行えた場合におけるユーザ情報の推測の「確信度」と，推測の被害を受けた「人数」と，推測によってうける「被害のレベル」の掛け算と考えた．なお,「被害のレベル」とは，推測によって，どのくらいの精神的・経済的被害を受けるかを示したものである[9]．そして,「被害が起こる可能性」を既知ユーザ攻撃などの推測が行える可能性と考え，ある属性がある事業者によって推測されてしまう推測リスクを以下のように定義した．

$$risk_{i,j} = \left(C_{i,j} \times U_j \times w_i\right) \times \left(P_j\right)$$

ここで，iは属性の種類を表し，jは推測を行う事業者を表す．$C_{i,j}$は事業者jにおいて属性iをどのくらいの「確信度」で推測できるかを表す．U_jは，その属性を，事業者jがどのくらいの「人数」分推測できたかを表す．w_iは，属性iが推測された場合における「被害のレベル」を表す．そして，P_jは，事業者jにおいて既知ユーザ攻撃による推測が行える「可能性」を表す．

また，各事業者の推測リスクの合計は，各事業者jで推測できる属性iの推測リスクとなるので，合計リスク値は以下の式で表せる．

$$Risk = \sum risk_{i,j}$$

ただし，j={最上位の属性保持事業者を除く属性保持事業者，配信事業者，ID 管理事業者}である．iは各属性事業者が持つユーザ情報の種類であるので，例えば，属性保持事業者 A が「年収」のユーザ情報を持ち，属性保持事業者 B が「好み」のユーザ

情報を持つ場合は，i={年収，好み・・・}のようになる．

以降の節に，「確信度」$C_{i,j}$，「人数」U_jの計算方法について説明する．なお，本論文ではリスク計算の簡易化のために，全てのw_iを 1 と置いて計算する．また，P_jは，シミュレーションを行って近似する．

5.1.1 $C_{i,j}$の計算方法

既知ユーザ攻撃によるユーザ情報の推測の確信度は，3.2 節で示した考え方で計算する．最初に，受け取ったユーザリストに既知ユーザが存在しなかった場合を考える．この場合，既知ユーザ攻撃は出来ないので，対象となるユーザ情報の値の分布から，ユーザ情報を推測することになる．例えば，年収が 500 万円台の人が全体で 10%という分布であれば，年収が 500 万円台であるという確信度は 10%となる．分布からのユーザ情報を推測の確信度を$_{Distribution}C_{i,j}$と置く．

次に，ID ルータを導入する前の，既提案のマッチング連係方式における，確信度の計算方法について説明する．この場合，3.2.3 節で説明したように，既知ユーザ攻撃を行う事業者である，ユーザリストを受け取る事業者は，直前の属性事業者保持するユーザ情報種類が解る．そのため，既知ユーザ攻撃を行う事業者は，ユーザ情報を確信 100%で推測可能である．

続いて，ID ルータによる改良版条件マッチング連係方式における，確信度の計算方法について，説明する．ID ルータによって，直前の属性事業者保持するユーザ情報種類が解らないので，確信度は 100%にならずに，直前の属性保持事業者の候補を考慮して確信度を求める．

まず，「単一既知ユーザ推測」の場合の確信度の計算方法を図 4 の例を用いて説明する．この例では，既知ユーザ攻撃を行う事業者である，ユーザリストを受け取る事業者以外に，属性保持事業者が 4 つ存在し，それぞれ「年収」「年齢」「好み」「現在地」を保持している．ここで，既知ユーザ攻撃を行う事業者が受け取ったユーザリストには，user1 という既知ユーザ 1 名が含まれており，「年収」が「400 万円台」で，「年齢」が「20 代」で，「好み」が「ゲーム」である．ここで，直前の属性保持事業者が「年収」を保持している事業者である可能性をpとおくと，逆に直前の属性保持事業者が「年齢」か「好み」か「現在地」を保持する事業者である可能性は$1-p$である．もし，直前の属性保持事業者が「年収」を保持している事業者であれば，user5 と user6 の「年収」は「400 万円台」であると推測でき，逆に直前の属性保持事業者が「年齢」か「好み」か「現在地」を保持する事業者であれば，既知ユーザ攻撃が出来なかった場合と同様に，分布からのユーザ情報を推測しかできない．よって，「単一既知ユーザ推測」によってユーザ情報が推測される場合の確信度は，$p \times 100\% + (1-p) \times {}_{Distribution}C_{i,j}$となる．

続いて，「合致既知ユーザ推測」の場合の確信度の計算方法を図 4 の例を用いて説明する．この場合は，既知ユーザ攻撃を行う事業者が受け取ったユーザリストには，user1 と user3 という 2 名の既知ユーザが含まれており，「年収」と「年齢」は一致しているが，「好み」は一致していない．そのため，直前の属性保持事業者は，「年収」か「年齢」か「現在地」を保持する事業者のいずれかである．ここでも先ほどと同様に，直前の属性保持事業者が「年収」を保持している事業者である可能性をpと置くと，逆に直前の属性保持事業者が「年齢」か「現在地」を保持する事業者である可能性は$1-p$である．よって，同様に，「合致既知ユーザ推測」によってユーザ情報が推測される場合の確信度も，$p \times 100\% + (1-p) \times {}_{Distribution}C_{i,j}$となる．

よって，既知ユーザ攻撃によるユーザ情報推測の確信度$C_{i,j}$は，ID ルータ導入前の既提案のマッチング連係方式と，ID ルータによる改良版条件マッチング連係方式に分けて考えられ，以下のような式で表せる．

$$C_{i,j} = \begin{cases} 100\% & (ID\text{ルータが無い場合}) \\ p_i \times 100\% + (1-p_i) \times {}_{\mathrm{Distribution}}\mathrm{C_{i,j}} & (ID\text{ルータが有る場合}) \end{cases}$$

但し，リスク計算の簡単のために，属性iにおけるp_iを以下のように定義する．

$$p_i = \begin{cases} 1/|\mathrm{Attr}| & (\text{「単一既知ユーザ推測」の場合}) \\ 1/(|\mathrm{Attr}| - |\mathrm{Attr_{unmatch}}|) & (\text{「合致既知ユーザ推測」で，} i \in \mathrm{Attr_{match}} \text{の場合}) \\ 0 & (\text{「合致既知ユーザ推測」で，} i \notin \mathrm{Attr_{match}} \text{の場合}) \end{cases}$$

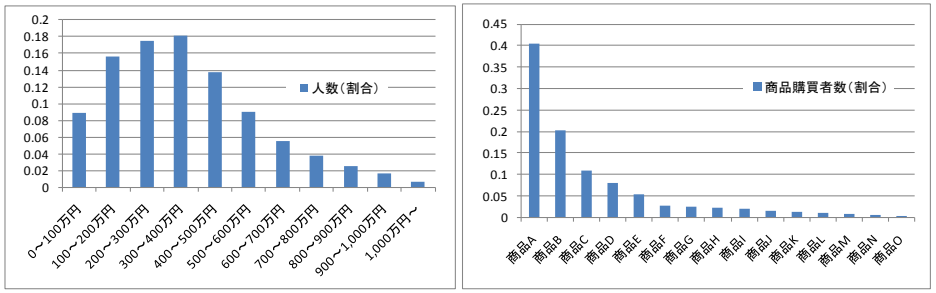
ここで，Attrは，全ユーザ情報の種類を表し，|Attr|は全ユーザ情報の種類数である．例えば，図 4 の例の場合は，属性保持事業者は「年種」か「年齢」か「好み」か「現在地」の 4 事業者であるので，$|\mathrm{Attr}| = 4$である．よって，図 4 の例での「単一既知ユーザ推測」の場合，「年収」と「年齢」と「好み」と「現在地」で$p_i = 1/4$となる．

また，$\mathrm{Attr_{unmatch}}$は，「合致既知ユーザ推測」における，合致しなかったユーザ情報種類を表し，$\mathrm{Attr_{match}}$は，合致したユーザ情報種類を表す．例えば，図 4 の「合致既知ユーザ推測」の例の場合は，合致したユーザ情報種類は「年収」と「年齢」であるので，$\mathrm{Attr_{match}}$={年収,年齢}となる．また，合致しなかった属性種類は，「好み」であるので，$\mathrm{Attr_{unmatch}}$={好み}となり，$|\mathrm{Attr_{unmatch}}| = 1$となる．よって，図 4 の例では，「年収」と「年齢」については$p_i = 1/(4-1) = 1/3$となり，「現在地」と「好み」では，$p_i = 0$となる．

5.1.2 U_jの計算方法

U_jは，既知ユーザ攻撃によって，事業者jがどのくらいの「人数」の属性を推測できたかを表す．事業者jが受け取ったユーザリスクには既知ユーザが含まれているため，事業者jがユーザ情報を推測できるユーザの「人数」は，この既知ユーザ以外のユーザの人数となる．よって，「受け取ったユーザリストの人数」を$_{Received}U_j$，「ユーザリスト中の既知ユーザ数」を$_{KnownUserInReceived}U_j$と置くと，以下の式で表せる．

$$U_j = {}_{Received}U_j - {}_{KnownUserInReceived}U_j$$



(a) 年収情報データのヒストグラム　(b) 商品購買者情報のヒストグラム

図 7 評価用データのヒストグラム

5.2 評価用のデータ

ユーザ情報の推測リスクを評価するために，以下の 2 種類のデータを用いた．

(a) 偏りが少ないユーザ情報：年収情報

(b) 偏りが激しいユーザ情報：商品購買者数情報

年収情報は，実際の年収分布[10]に従ったデータを生成した．商品購買者数情報は，全体の上位 2 割の商品が 8 割のユーザによって買われているような分布でデータを生成した．これらのデータのヒストグラムを図 7 示す．これらの分布に従って，100,000 人分のユーザ情報を生成して評価を行った．

5.3 単一の属性保持事業者に対するユーザ情報の推測リスクの評価

まずは，単一の属性保持事業者に対して ID ルータ機能を用いた改良版条件マッチング連係方式の評価を行い，局所的なユーザ情報の推測リスク(*risk*)の低減を評価する．

5.3.1 評価方法

単一の属性保持事業者に対するユーザ情報の推測リスクの評価では，年収情報を持つ属性保持事業者と，商品購買者数情報のデータを持つ属性保持事業者と，配信事業者との構成で評価を行う．そして，どちらか片方の属性保持事業者だけで条件マッチングを行い，その後，配信事業者が結果のユーザリストを受信するという動作になる．また，配信事業者が，属性保持事業者が保持するユーザ情報を推測するという想定で評価を行う．

評価では，年収情報と商品購買者数情報のデータをそれぞれ用いて，属性保持事業者におけるリスク値を算出した．シミュレーションは 50 名の既知ユーザをランダムに選んで 10,000 回行い，結果はこれらの平均を取っている．

5.3.2 評価結果

まず，ID ルータ機能の導入による効果を調べるために，ID ルータ機能無しの条件マッチング連係方式と，ID ルータ機能有りの改良版条件マッチング連係方式における，

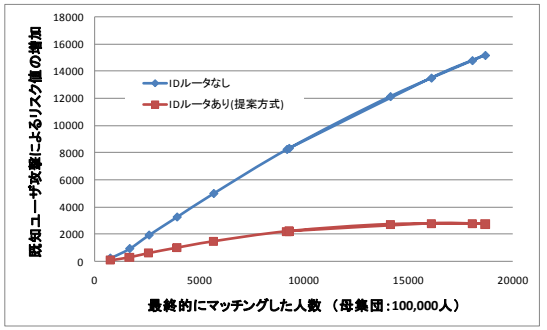


図 8 年収情報の各種条件に応じたリスク値の増加

既知ユーザ攻撃によるリスク値の増加について調べた．図 8 に結果を示す．これは，年収情報について，それぞれの検索条件を選んでシミュレーションしたものである．つまり，例えば「年収＝300～400 万円台」という検索条件が選ばれた場合は，最終的に配信された人数は 100,000 人の約 18%となる 18,000 人となり，その場合のリスク値の増加をプロットしている．なお，プロットしている値は，既知ユーザ攻撃によるユーザ情報の推測リスクから，既知ユーザ攻撃が行われなかった場合のユーザ情報の分布情報からの推測リスクを引いた，推測リスクの差分である．

図 8 に示したように，ID ルータ機能無しの条件マッチング連係方式の場合は，最終的にマッチングしたユーザ数が多ければ多いほど，リスク値が増加している．それに対し，ID ルータ機能有りの改良版条件マッチング連係方式の場合は，増加が非常に緩やかである．これは，直前の配信事業者が解る場合は，直前の事業者のユーザ情報の確信度が 100%で解るため，ほぼ単調にリスク値が増加するが，直前の配信事業者が解らない場合は，直前の属性保持事業者の候補数や，設定されたマッチング条件のユーザ情報の分布に応じて，確信度が抑えられるため，増加が抑えられるからである．

また，図 9 に，既知ユーザ数を 5～60 人まで変化させた場合における，リスク値の変化を示す．このグラフは，商品購買者情報のデータを用いて，「購買商品＝商品 C」(約 10%の約 10,000 人がマッチング)という条件を選択し，既知ユーザを 5,10,20,30,40,50,60 人にした場合におけるリスク値の増加を示したものである．このグラフが示すように，ID ルータ機能有りの改良版条件マッチング連係方式は，ID ルータ機能が無い既提案方式の条件マッチング連係方式に比べてリスク値が小さく，また，たとえ，既知ユーザが増えたとしても，リスク値が大幅に増加することは無いことが解る．

以上の結果から，ID ルータ機能による改良版条件マッチング連係は，ID ルータ機能が無い既提案方式の条件マッチング連係方式よりも，ユーザ情報のリスクを低減で

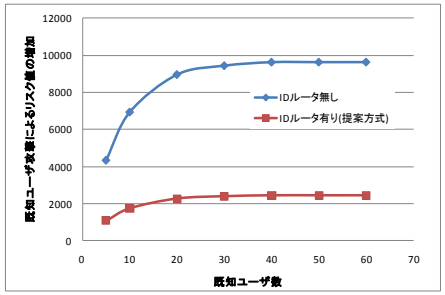


図 9 既知ユーザ数に対する
リスク値(risk)の増加

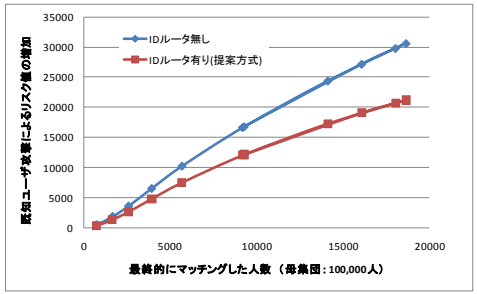


図 10 各事業者でのユーザ情報の
推測リスク値の合計(Risk)の増加

きることができることが解った．

5.4 複数の事業者に対するユーザ情報の推測リスクの評価

次に，条件マッチング連係方式の各事業者に対するリスク値を評価する．先ほどの評価とは違い，5.1 節で説明した合計リスク値($Risk$)を求め，ID ルータ機能無しの条件マッチング連係方式と，ID ルータ機能有りの改良版条件マッチング連係を比較する．

5.4.1 評価方法

この評価では，年収情報を持つ属性保持事業者 A と，商品購買者数情報を持つ属性保持事業者 B の 2 つの属性保持事業者で条件マッチングを行った場合を想定して評価を行う．条件マッチングの順番は，まず，属性保持事業者 A で年収についての条件でマッチングを行い，その後，属性保持事業者 B で商品購買者数についての条件でマッチングを行うという順序である．また，属性保持事業者として属性保持事業者 C も存在するものとして，全体で属性保持事業者は 3 事業者あるとする．

この評価では，属性保持事業者 B に設定する条件を図 7 (a)で示した商品購買者数データの「購買商品＝商品 F」（約 3%のユーザが合致する）と固定的に設定し，属性保持事業者 A に設定する条件を，図 7(b)で示した，年収データの検索条件を「100 万円以下」から「1500 万円以上」までそれぞれ個別に設定して評価を行った．値は，50 名の既知ユーザをランダムに選んで，シミュレーションを 10,000 回行った平均である．

5.4.2 評価結果

図 10 にシミュレーション結果を示す．このグラフに示したように，条件マッチング連係のシステム全体においても，ID ルータ機能有りの改良版条件マッチング連係方式の方が，リスク値を小さくできる．

以上の結果から，単一の属性保持事業者における推測リスクの評価と同様に，ID ルータ機能を導入した改良版条件マッチング連係方式によって，システム全体の推測リスクが減ることが解った．

6. まとめ

本論文では，以前提案した条件マッチング連係方式に，既知ユーザ攻撃が存在する事を述べ，既知ユーザ攻撃による「単一既知ユーザ推測」と「合致既知ユーザ推測」という推測方法を明確にし，既知ユーザ攻撃によるユーザ情報の推測リスクを低減する方式として ID ルータ機能を用いた改良版条件マッチング連係方式を提案した．この方式を用いることで，ユーザ情報の推測リスクを低減できることを示した．

また，本論文では，リスク評価の簡便化のために，各種ユーザ情報の重要度などは考慮しなかった．しかし，実際はユーザ情報の種類に応じて，ユーザ情報の漏洩の被害の度合いが違うので，実際のリスクを考える際は，重要度なども考慮してリスクの計算式を立てる必要があるだろう．

また，本論文では，直前の属性保持事業者のユーザ情報の推測リスクのみを検討していた．今後は，直前に限らず，上位の全ての属性保持事業者のユーザ情報の推測リスクも考慮したいと考えている．

参考文献

[1] 産業競争力懇談会(Council on Competitiveness-Nippon): 個人情報や企業情報を安全に活用するためのクラウドコンピューティング基盤の整備, 産業競争力懇談会 2010 年度 プロジェクト最終報告, 2011. http://cocn.jp/common/pdf/thema30.pdf

[2] Andrew C. Yao: Protocols for secure computations, In Proc. 23rd. IEEE Symposium on the Foundations of Computer Science (FOCS), pages 160–164. IEEE, 1982.

[3] 野島良，門林雄基: 適度に安全な秘匿共通集合計算プロトコルのパフォーマンスについて，暗号と情報セキュリティシンポジウム(SCIS2008), 3E1-2.

[4] 竹之内隆夫，豊田由起，南澤岳明: ユーザ情報漏洩リスクを軽減した条件マッチング連係，電子情報通信学会 2010 年総合大会論文集 B-7-98.

[5] 豊田由起，竹之内隆夫，南澤岳明: 条件マッチング連係におけるユーザ情報漏洩リスク軽減の改善手法，電子情報通信学会 2010 年総合大会論文集 B-7-99.

[6] OASIS Standard: Assertions and Protocols for the OASIS Security Assertion Markup Language(SAML) V2.0, 2005. http://docs.oasis-open.org/security/saml/v2.0/saml-core-2.0-os.pdf.

[7] OpenID Foundation: OpenID Authentication 2.0 - Final, 2007. http://openid.net/specs/openid-authentication-2_0.html.

[8] 日本情報経済社会推進協会(JIPDEC): ISMS ユーザーズガイド-JIS Q 27001:2006(ISO/IEC 27001:2005)対応- -リスクマネジメント編-, 2008

[9] 日本ネットワークセキュリティ協会: 情報セキュリティインシデントに関する調査報告書 第 1.1 版, 2010. http://www.jnsa.org/result/incident/data/2009incident_survey_v1.1.pdf

[10] 国税庁 長官官房企画課:平成２１年分 民間給与実態統計調査, 2010. http://www.nta.go.jp/kohyo/tokei/kokuzeicho/minkan2009/pdf/000.pdf