社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2004－ＣＳＥＣ－26　(42)
2004／7／21

# 情報エントロピーを用いたプライバシ保護手法：LooM

今田　美幸†　　高杉　耕一†　　太田　昌克†

† NTT 未来ねっと研究所，〒 180-8585　東京都武蔵野市緑町 3-9-11

E-mail: †{imada,ohta}@ma.onlab.ntt.co.jp, ††takasugi.koichi@lab.ntt.co.jp

**あらまし**　本稿では，ユビキタスネットワーキング環境において，プライバシ保護と利便性を適度なバランスで保つことを目指したプライバシ保護のための提案手法 LooM（Loosely managed privacy protection Method）について述べる．LooM は，従来の 1 対 1 通信を基本としたプライバシ保護をユビキタスネットワーキング環境に適用した場合の利便性低下に対する問題解決法として，プライバシ情報をデータベースとして管理し，匿名性の評価尺度として情報エントロピーを用いる手法である．人の移動に伴いデータベースのユーザ集合が変更しても単一量で表現できるような関数をを提案し，シミュレーションにより有効性を確認する．また，LooM は，追跡攻撃やプライバシ情報からのユーザ ID 推定を回避できることを示す．

**キーワード**　情報エントロピー，プライバシ保護，ユビキタスネットワーキング環境，決定木

# A Privacy Protection Method using Information Entropy: LooM

Miyuki IMADA†, Koichi TAKASUGI†, and Masakatsu OHTA†

† NTT Network Innovation labs., Midori-cho 3–9–11, Musashino-shi, Tokyo, 180-8585 Japan

E-mail: †{imada,ohta}@ma.onlab.ntt.co.jp, ††takasugi.koichi@lab.ntt.co.jp

**Abstract**　We propose a privacy protection method for use by service providers supplying users with services where user information requires privacy in the ubiquitous networking environments. The main feature is the use of an entropy value as a criteria for the evaluation of anonymity. This prevents loss of the convenience of services in the target environment. Objects in the ubiquitous environments will include more private information. This is because information gathered by sensors and so on will be added to static information such as sex and age. Furthermore the end-to-end communication assurance that are conventionally used become more complicated. In the new method, privacy information is managed in a database on a server. The novel idea behind the method is that a calculated by using this database decides whether or not the provider is given access to private information. The entropy value is independent of the number of people listed in the database, so the values are consistent over time. We implement the proposed method and evaluate it though simulations, thus confirming its anonymity feasibility.

**Key words**　information entropy, privacy protection, ubiquitous networking environment, decision tree

## 1. はじめに

将来のユビキタスネットワーキング環境においては，センサや携帯端末のような多種多様な端末が遍在化し，互いに協調連携しながら，ユーザの嗜好や環境に適応したアプリケーションサービスが広域で利用提供できるようになると考えられる．そのような適応型サービスを提供するためには，個人のプライバシ情報を安全かつ便利に利用できる仕組みが必要不可欠である．

現在市場には，アンケートや会員登録によりユーザが提供するプライバシ情報と引き換えに，ユーザはサービスプロバイダ（以下プロバイダと呼ぶ）から景品やポイントカードによる特典を受けるというサービスがある．プロバイダは，誰が，いつ，何を購入したかなどの顧客の購買傾向を分析するためのデータとしてそのプライバシ情報を用い，結果は主に商品配置，品揃え，キャンペーン展開の戦略を立てるために活用している．このように，お金の代わりに情報を提供し，その対価として得るサービスを“Give and Take 型サービス”と呼ぶ．ユーザは特典を得るために安易にプライバシ情報を提供してしまう傾向があるが，提供したプライバシ情報の管理はプロバイダに一任されている場合が多く，その漏洩が社会問題になっている．

ユビキタスネットワーキング環境で想定されるサービスも，Give and Take 型サービスの延長線上にあると考えられる．違いは，プロバイダのプライバシ情報の取得方法，ユーザへの特典である．まず，取得方法であるが，これまではアンケートや会員登録という手段を用いていたが，ユビキタスな環境ではセンサやGPSなどの手段を用いる．ユーザへの特典は，これまでは景品やポイントカードであったが，ユビキタスな環境ではユーザの置かれている状況や嗜好にあったサービスを提案したり提供することになる．これまでは，プロバイダは利用したいプライバシ情報をユーザの許可を得た上で個別に取得していた．しかし，センサなどを用いて取得したプライバシ情報に対しても同じ手法を適用すると，現状にはない新たな問題が発生する．1つは，公共の場所に設置されているセンサを用いて，ユーザの了解なしに採取・利用されるプライバシ情報に対する保護に対する問題がある．また，センサから取得する情報をサービス提供に利用する際，ユーザに個別に利用許可を求める従来手法だと利便性が低下するという問題もある．

本稿では，環境中にいる全てのユーザのプライバシ情報をネットワーク上のサーバでデータベース（DB）として管理し，情報エントロピーを用いて匿名性を数値で表現する LooM(Loosely managed privacy protection Method) を提案する．LooM は，システムで定めた匿名性の基準値を元に，プロバイダに対してどこまでプライバシ情報を教えられるかを情報エントロピーを用いて表現する．LooM は，プライバシ保護と利便性のバランスを適度に保つことを目標としている．2章では，想定するプライバシ保護として，プライバシ情報の定義，想定するプロバイダのサービス，ターゲットとするプライバシ保護，機密性との関係について述べる．3章では，本稿で提案する LooM の説明として，ユビキタスネットワーキング環境におけるプライバシ保護をモデル化し，その適正について評価する．4章では，LooM のシステム化として LooM サーバの概要とプロトタイプ実装について述べる．5章では，考察として，LooM への攻撃とその回避方法，関連研究について述べ，最後にまとめを述べる．

## 2. 想定するプライバシ保護

### 2.1 プライバシ情報の定義

プライバシ情報とは，ユーザを表現する情報であると定義する．ただし，システムがユーザを一意に識別するためのユーザ ID はプライバシ情報に含めないとする．

ユビキタスネットワーキング環境で扱うプライバシ情報には，ユーザが端末から入力した静的な情報と，ユーザの挙動や振る舞いをシステムが感知して入手する動的な情報がある．前者を静的プライバシ情報 (S-privacy：Static privacy)，後者を動的プライバシ情報 (D-privacy：Dynamic privacy) と定義する．S-privacy には，名前，住所，年齢，性別，病歴，家族構成などのユーザの属性，嗜好，意図，およびこれらの履歴などが含まれる．D-privacy には，位置，生体，温度，人検知などのセンサによって取得できる情報，時刻，電子マネーでの購入記録，乗車記録などのユーザの行動に付随して取得できる情報，およびこれらの履歴などが含まれる．

### 2.2 想定するプロバイダのサービス

我々は，ユビキタスネットワーキング環境で，プロバイダから次のようなサービスが提供されると想定している．図1に示すように，ユビキタスネットワーキング環境は，複数の端末，センサ，ネットワーク，サーバから成ると考えられる．ネットワークには，インターネットの他に，ホームネットワーク，センサ

ネットワーク，アドホックネットワークといったような様々な種類があり，それらはゲートウェイなどを介して相互接続されている．ネットワーク上には，幾つかのサーバが配置してあり，ロケーション管理，コンテキスト管理，センス情報管理，認証，攻撃防御のための監視，攻撃者の追跡・特定などを行っている．また，複数の固定端末や携帯端末がネットワークに接続され，端末を経由してユーザやプロバイダがサービスを利用・提供する．

ユーザが保持している携帯端末は，ユーザ自身が設定したS-privacy，GPSとの連携によるユーザの位置情報などを保持している．携帯端末は，ユーザの移動に伴い，移動先のネットワークに適宜接続する．

例として，移動を伴うサービスを考える．携帯端末を所持したユーザが，家から町のレストランに移動する．この場合，携帯端末は当初接続されていたネットワークから切り離され，レストラン到着時に再度ネットワークに接続される．ユーザは，レストランに入り，ディスプレイ付きテーブルに座る．この様子をGPSやセンサが検出し，位置情報やセンス情報としてサーバに送信する．センスサーバは，このネットワークに接続しているコンテンツ提供サービスを行うプロバイダに対し，これらの情報を通知する．プロバイダは，ユーザの携帯端末から嗜好情報を取得し，ユーザの好みのコンテンツをディスプレイ付きテーブルに送信する．このようにして，ユーザは食事が出てくるまでの間，好みのコンテンツを見て過ごすことができる．

### 2.3 ターゲットとするプライバシ保護

プライバシ保護について，プライバシ情報から実空間の個人を特定する問題について考える必要がある．プライバシ情報を元に実空間での個人を特定する場合，仮想空間でプライバシ情報を用いたユーザIDの推測・特定を行い（第1ステップ），次に実空間上でそのIDをもつユーザを特定する（第2ステップ）という2段階のステップを踏む．ただし，仮想空間上でのプライバシ情報からユーザIDを対応づけることと，仮想空間のユーザIDと実空間でのユーザ本人を対応づけることは，別のステップである．

2.2節において，プロバイダがサービス提供に使う情報は，レストランID，ディスプレイ付きテーブルID，人が座ったというセンス情報，ユーザの嗜好情報であって，ユーザIDは必要としていない．第1ステップとは，これらのユーザのプライバシ情報を元にプロバイダが仮想空間上のユーザIDを推測・特定することである．

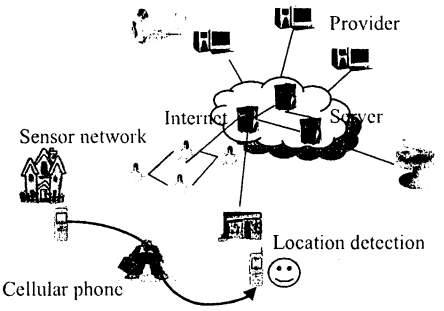



図1 想定するユビキタスネットワーキング環境とサービス
Fig. 1 Supposition of networking environments and services

第2ステップでは，ユーザIDから実空間のユーザを特定するのであるが，このとき，プロバイダが実空間のユーザ集合に対して背景知識があるか否かにより，個人特定が可能かが変わってくる．2.2節の例においては，偶然，プロバイダがユーザと同じレストランにいて，レストランにいる人がどんな人かという情報をプロバイダが背景知識として持っているとする．プロバイダは，自分が提供したコンテンツを見ているユーザが自分の目で特定できるような場合，ユーザIDと実空間上の個人がマッピングできる．また，別の例として，日本人であれば首相が一人しかいないという背景知識を持っている．「日本の首相」からユーザIDが分かった場合，背景知識とユーザIDとのマッピングで実空間上の人物が誰であるかを特定できる．このように，ユーザ集合に対して背景知識がある場合，背景知識とユーザIDのマッピングにより個人特定が可能と考える．しかし，ユーザ集合に対して背景知識がない場合，第1ステップで仮想空間のユーザIDが推測困難であれば，結果的に実空間の個人特定も困難になると考えられる．よって，本稿では，このポイントについて述べる．

### 2.4 機密性とプライバシ情報の関係

プライバシ保護を考える場合，対象とするプライバシ情報の機密性の高さによって保護方法が変わるため，ユビキタスネットワーキング環境で扱うプライバシ情報がどの程度の機密性を持っているかを整理する必要がある．以下に，機密性の高さと保護方法との関係を述べる．

機密性が高い情報とは，財産情報やクレジットカード番号ような攻撃によりユーザ個人へ深刻な危害が

及ぶ情報である．機密性が低い情報とは，「今ドアの前に立っている」といったような攻撃されてもユーザ個人にそれほど深刻な危害が及ばない情報である．

明らかに機密性が高い情報に関しては，耐タンパなデバイスからなるICカードなどを用いてプライバシ情報の格納・管理し，本情報を利用する際は従来から行われていたようなエンドエンドでの認証，本人証明，アクセス制御を行い，暗号を用いて通信路での盗聴や改竄を防ぐ必要がある．また，明らかに機密性の低い情報は，ほとんどの場合，保護対策が不要である．

しかし，2.2節で示したようなサービスで使うプライバシ情報の多くは，このいずれでもない．2.2節で使っているプライバシ情報は，ユーザの嗜好や行動履歴であり，ある程度の匿名性が確保が必要な情報である．これを機密性中程度のプライバシ情報と定義する．

本稿では，中程度の機密性に関してなんらかの定量的な評価尺度を与える手法について示す．

## 3. LooM

### 3.1 ユビキタスネットワーキング環境におけるプライバシ保護モデル

#### 3.1.1 エントロピーと匿名性評価尺度

ユーザ集合において，プライバシ情報の与え方によってユーザIDの特定のし易さが決まる．この情報の与え方は，決定木学習の分類アルゴリズムと同じ考え方である．決定木学習では，効率的に分類をするために，情報を受け取る前後の情報利得を計算し，情報利得の多い情報を決定木のノードとして，より少ないノードで決定木を作る．つまり，より少ない情報でユーザIDを特定すること目指した分類アルゴリズムである．逆に言うと，情報利得の少ない情報を決定木ノードをすれば，ユーザIDは特定されにくくなる．この考えに基づき，LooMでは，情報利得が少ないプライバシ情報を決定木のノードとして選び，結果的に匿名性を確保する．

情報利得の計算には，「情報エントロピー（以下エントロピーと略す）[7]」を用いている．エントロピーとは，ある情報の集合である情報源に対して，その要素である情報を受け取る前後の不確かさの変化を定量的に表す量である．エントロピーは，あるプライバシ情報を受け取ることによって得られる情報量である．例えば，50人の男性と女性から成る100人のユーザの集合を考える．最もエントロピーが高いのは，「性別」に関する情報のような全てが同じ確率で生じる情報を与えられた場合であり，この時最も匿名性が高いと定義する．逆に最も匿名性が低い場合は，不確定さがない発生確率が1の特定のプライバシ情報を与えた場合であり，この時エントロピーが最小になる．つまり，100人の中でAliceという名前の人が1人しかいない場合，Aliceという名前を情報として受け取ることで，不確かさがなくなり，エントロピーが最小になる．エントロピーは，匿名にしたい対象がAliceという1人のユーザであっても，（血液型，A型）というユーザ集合であっても同じ手法で計算できるため，匿名にしたい対象を自由に変更できる．

表1 プライバシ情報の管理形式
Table 1 Management format of privacy information

| ユーザID | S-privacy | D-privacy |
|---|---|---|
| *Alice* | (性別, 女性), (年齢, 6歳), (職業, 小学生) | (朝の体温, 36.5), (8時の居場所, 家) |
| *Bob* | (性別, 男性), (年齢, 12歳), (趣味, 水泳) | (朝の体温, 36.9), (10時の居場所, 公園) |

本稿では，ユーザのプライバシ情報を情報源とし，プライバシ情報を受け取る前後でのユーザIDの特定に関する不確かさの減少量をエントロピーで表現し，匿名性の定量化を実現する．プライバシ情報は，ユーザID毎に（属性，属性値）のペアとして，LooMサーバのDBで安全に管理する．属性とは，性別や年齢といったような個人を特徴付ける情報種別を表し，属性値とは，女性や10歳といったような属性に対応する値を表す．（表1参照）

#### 3.1.2 ユーザ集合の違いとエントロピー値

2.章で示したように，ユビキタスネットワーキング環境では，ネットワークに接続されるユーザ数が時間と共に変化する．よって，ユビキタスネットワーキング環境におけるプライバシ保護にエントロピーの概念を適用する場合，エントロピー計算の前提であるユーザの集合が常に一定という条件が定常的に成り立たない．よって，システム化に向けては，ユーザ集合に依存しない匿名性の定量化が必要になる．そこで，ユーザ集合の変化が無視できる$\Delta t$時間単位にエントロピー値を計算し，エントロピー値を正規化することでユーザ集合の違いを吸収する．本稿では，正規化した値を匿名性の定量的評価尺度に用いることを提案する．

正規化の方式には，様々な方式が考えられるが，本稿では，以下のような方式を採用する．保護したい属

性において，匿名にしたいユーザと同じ属性値のユーザを正，そうでない属性値を負という2つのクラスに別ける．全てユーザ集合における匿名にしたいユーザの正負に関するエントロピーを $H_0$，ある属性Fが判明した時のエントロピーを $H_F$ とした場合に，Fを知ることにより減少するエントロピーの割合を $D_F$ とし，これを匿名性の評価尺度に用いる．

$$D_F = \frac{H_F}{H_0}$$

$$H_0 = -P_+ \log P_+ - P_- \log P_-$$

$$H_F = \sum_{i=1}^{n} P_i(-P_{i+} \log P_{i+} - P_{i-} \log P_{i-})$$

$P_+$：全てのユーザ集合において，正である確率
$P_-$：全てのユーザ集合において，負である確率
$P_{i+}$：Fの属性値の集合を $\{\nu_1, \nu_2, \nu_3, \nu_4, \cdots, \nu_n\}$ とした場合，Fが $\nu_i$ である確率（Fにおける正である確率）

$P_{i-}$：Fの属性値の集合を $\{\nu_1, \nu_2, \nu_3, \nu_4, \cdots, \nu_n\}$ とした場合，Fが $\nu_i$ でない確率（Fにおける負の確率）

$D_F$ の閾値は，適用するアプリケーションのセキュリティレベルに応じて規定する．

### 3.2 $D_F$ の評価

実際の人のプライバシ情報を模擬したDBを用いて，ある時間Δtのユーザ集合におけるユーザAliceのID（以降Alice IDと記す）の匿名性について評価を行った．

プロバイダがサービスを起動する際に利用したい属性の数を指定属性数と呼ぶ．プロバイダがDBの中から（属性，属性値）をいくつ指定したかを指定属性数とし，指定属性数と $D_F$ の関係について検証した．DBは，10種類の属性から成るランダムサンプルとし，属性値の分布は均一になるようにした．属性毎の全ての組み合わせについて，$D_F$ 値を求め，その平均を出した．本評価では，ユーザ数が10人と100人の場合について比較を行った．10人の中にも100人の中にも，Alice IDが含まれている．

結果を図2に示す．$D_F$ 値は，10人と100人共に，指定属性数が多くなるにしたがって小さくなる．つまり，指定する属性が多くなると，Alice IDを特定し易くなり，匿名性が低下することが言える．

10人と100人を比較した場合，10人の方が $D_F$ の減少率が高い．これは，ユーザ数が少ない場合，少ない属性指定でAlice IDが特定でき，ユーザ数が多い場合，属性を多く指定してもAlice IDは特定し難いことを意味する．よって，ランダムサンプルを前提とした場合，ユーザ数が多いほどプロバイダが利用できる<属性，属性値>が多い．

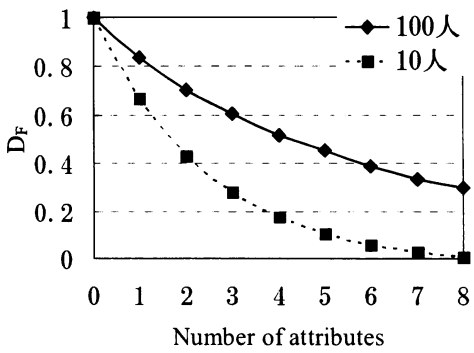



図2 $D_F$ と指定属性数の関係
Fig. 2 Relation between $D_F$ and Number of attributes

$D_F$ の具体例として，$D_F$ とユーザ数の関係を示す．$D_F$ の閾値を0.4とした場合，ユーザ数10人の場合，プロバイダは属性を1つまたは2つしか指定できない．この時，Aliceは2人に1人に絞られてる．ユーザ数100人の場合は，1〜6個の属性を指定でき，この時，Aliceは4人に1人に絞られてる．

上記より，$D_F$ は，ユーザ数が異なる場合でも同一尺度で匿名性を評価できる傾向を示すことを確認した．実用システムとしての $D_F$ 値は，システムやアプリケーションの要求するセキュリティレベルによって異なるので，フィールド実験などを行うことでΔtと併せて今後詳細な評価を行っていく．

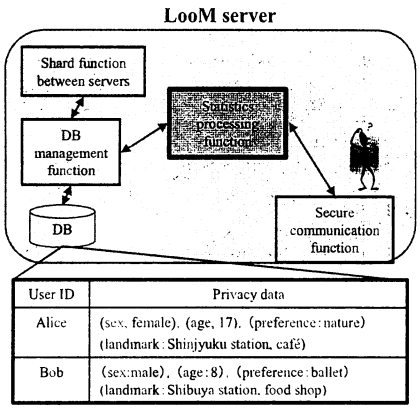
本評価では，ユーザ数が少ないところで検証を行ったが，現実的には要素数が3桁〜4桁以上になると想定され，適正な閾値の設定により実用上十分な匿名性が確保可能と考える．

## 4. LooMサーバ

### 4.1 概　　要

LooMサーバは，LooMの実現形式の一つであり，Δt毎の $D_F$ 値の計算と匿名性の閾値に応じてプロバイダに開示できるプライバシ情報を決定する．LooMサーバは，図1に示すネットワーク上に分散配置する．

図3にLooMサーバの構成を示す．LooMサーバは，プライバシ情報の登録削除を行うDB管理機能，LooMに基づき匿名性を評価する統計処理機能，ネットワーク上に分散配置されているLooMサーバ間でプライバシ情報の共有をするサーバ間共有機能，セキュアエージェントなどを用いてエンドエンドで高い安全性を確保したい場合に用いるセキュア通信支




| User ID | Privacy data |
|---|---|
| Alice | (sex, female), (age, 17), (preference: nature) (landmark: Shinjyuku station, café) |
| Bob | (sex:male), (age:8), (preference: ballet) (landmark: Shibuya station, food shop) |

図3 LooMサーバの構成
Fig. 3 Structure of LooM server

援機能[1]，プライバシ情報を格納するDB，およびDB管理機能から成る．

DB管理機能は，あらかじめTTL（Time To Live）で定めた範囲内のユーザのプライバシ情報をDBに登録し，ユーザのネットワークからの離脱を契機に削除する．

統計処理機能は，DBに格納されたプライバシ情報から$D_F$の計算を行う．システムのセキュリティポリシなどからあらかじめ$D_F$の閾値kを決定しておく．プロバイダが指定した属性から$D_F$を計算し，その値がkより大きい場合規定の匿名性が確保されているとみなし，プロバイダにその属性を持っているユーザが居ることを通知する．kより小さい場合，システムで定めた匿名性が確保できないとみなし，プロバイダには，その属性を持っているユーザの有無についての情報を示さない．プロバイダが指定した属性がDBにない場合も，ユーザの有無についての情報を示さない．

サーバ間共有機能は，DBのスケーラビリティ確保のために，分散配置したLooMサーバで管理しているプライバシ情報DBを共有して，仮想的な1つのDBとし，スケーラブルなプライバシ保護を考慮したサービスを提供する．

サーバに登録するプライバシ情報の機密性の高さの判断はユーザ自身しかできないため，LooMサーバに登録するプライバシ情報はユーザが決める．

### 4.2 属性数と計算量

LooMは，全ての属性の組み合わせについて$D_F$を計算している．このため，DBに登録する属性数が多い場合，計算量が増加するため，提供できる属性の組がなにかを判断するのに時間がかかる．計算許容時間から決定した属性数を閾値をxとし，Δtに管理するユーザのプライバシ情報の属性数がxより大きい場合，$D_F$の計算時間がシステム要件を満たさないと判断し，プロバイダが要求する属性の組のみ$D_F$値を求めるというQ & A方式を採る（以下Q & Aパターンと呼ぶ）．プライバシ情報の属性数がxより小さい場合，$D_F$の計算時間がシステム要件を満たすと判断し，全てのプライバシ情報について$D_F$を計算する（以下全計算パターンと呼ぶ）．

### 4.3 プロトタイプ

フィージビリティ確認のために，Windows XPの上でjavaでLooMサーバのDB管理機能，DB，統計処理機能をプロトタイプとして実装した．DB管理機能として，オープンソースのMySQL[8]を用いた．Δt内にユーザが登録したプライバシ情報の属性は，10種類とした．S-privacyの属性は，性別，職業，学校種別，血液型，D-privacyの属性は，朝の体温，8時～12時の1時間毎の居場所とした．匿名性を表す$D_F$の閾値kは0.5，ユーザ数は10人とした．

プロトタイプでは，4.2節で述べた2つのパターン，つまりQ & A方式でプロバイダから提示を求められた属性の組のみ$D_F$を計算し，提示の可否を決めるQ & Aパターン，あらかじめ提示可能な属性の組を計算しておく全計算パターンを実装した．匿名にしたいのはAlice IDであり，Aliceは性別が女性で，血液型はA型である．Q & Aパターンにおいて，Aliceと同じ性別と血液型を指定した場合を図4に示す．Q & Aパターンでは，閾値k以上ならば，「その属性を持った人は結構います」と返答し，閾値以下およびその属性をもった人がいない場合，「その属性を持ったひとはほとんどいません」と表示する．DBでは10人中5人が女性，かつA型の人は3名であり，その中の一人がAliceである．LooMでは，入力情報として与えられた（性別，女性），（血液型，A型）を元に，性別と血液型の属性から$D_F$を計算する．その結果，$D_F$=0.58となり，閾値k = 0.5より大きい値なので，LooMサーバは，「その属性を持った人は結構います」と返答する．つまり，LooMサーバは，性別と血液型に関するプライバシ情報をプロバイダに開示したことになる．

図5では，全ての$D_F$を求める全計算パターンを示す．これらは，$D_F > 0.5$の全ての属性の組みである．例えば，属性1つなら，性別，職業など，属性2つなら，職業と学校種別の組み合わせ，属性3つなら職業と学校種別と8時にいる場所といったような

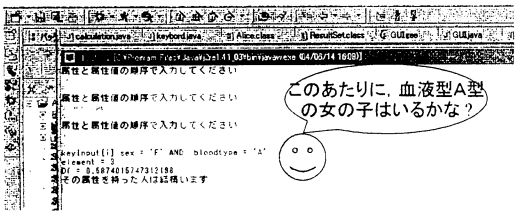



図4　Q & A パターン
Fig. 4　Case of Q and A

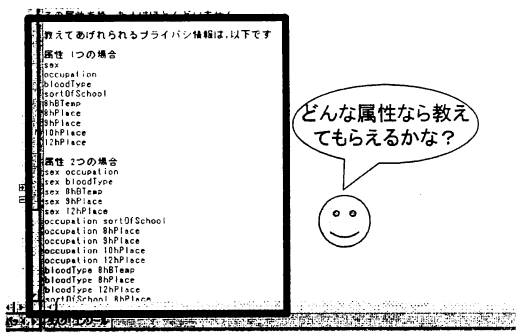



図5　全計算パターン
Fig. 5　Case of all answers

形式で提示する．$D_F$≦0.5の組み合わせについては，匿名性要件を満たしていないので提示しない．

## 5. 考　　察

### 5.1 LooMに対する攻撃

#### 5.1.1 追跡攻撃

ユビキタスネットワーキング環境では，ユーザの移動によりDBのユーザ集合が変わる．よって，Δtの前後でDBで管理する属性情報も異なる．LooMでは，ユーザ集合の変化が無視できるΔt毎に$D_F$を計算し直し，閾値を元にプロイバダへ提示できる属性の組を決めているため，プロバイダに提示できる属性の組もΔtの前後で変わり，Δtを跨がる追跡攻撃は回避できると考える．

次に，悪意のあるプロバイダが一人のユーザの行動を仮想空間上で追跡するような場合を考える．例えば，2.2節のサービスにおいて，ユーザが家からLooMサーバ管理下のレストランに移動したことを追跡している悪意のあるプロバイダがいたとする．LooMサーバでは，管理対象のユーザIDの匿名性が確保できないような属性情報をプロバイダに提示しない上，Δt毎にDBを更新するため，追跡中のユーザがいつLooMサーバの管理下から外れたかの識別が困難である．よって，LooMサーバの管理下の入る前後で，ユーザIDの追跡は困難になるため，追跡攻撃が回避できると考える．

#### 5.1.2 プライバシ情報蓄積によるユーザID特定

悪意のあるプロバイダがΔt内の属性の組を蓄積して，ユーザIDを特定するという攻撃が考えられる．対策としては，LooMがΔt内にプロバイダに開示した属性を蓄積しておき，その累積属性の組に対して$D_F$値を求め，閾値k以上なら，プライバシ情報を提示することで対処する．

悪意のあるプロバイダが，$D_F$値の閾値以下の属性の組だけを集めてユーザIDを推定するという攻撃も考えられる．つまり，4.3節で示したQ & Aパターンを例にとると，LooMが「その属性を持った人はほとんどいません」と答えた属性の組のみ収集して，ユーザIDを特定しようとする攻撃である．LooMでは，このように返答した場合，規定の匿名性を確保できない場合と，指定された属性が存在しない場合がある．つまり，LooMでは，DBに指定された属性がない場合にも，「その属性はありません」とは答えず，匿名性が確保できない場合と同じ答え方をする．よって，閾値以下の属性の組のみを集めたとしても，その中にはそもそも存在しないユーザの属性情報が混在しているので，悪意のあるプロバイダはDBを正しく再現できない．

これらのことより，閾値以上の属性情報を蓄積した場合も，閾値以下の属性の組を蓄積したとしても，ユーザIDの特定は困難であるため，本攻撃は回避できると考える．

#### 5.1.3 DBの拡張によるユーザID特定

LooMサーバのDBに登録してあるユーザのプライバシ情報がS-privacyであり，ユーザの移動がΔt前後でほとんどない場合，ユーザIDを特定されてしまう場合がある．そこで，サーバ間共有機能を用いて管理するDBを拡張し，Δt毎のDB変動の大きいD-privacyを含むDBを仮想的な1つのDBとすることで，Δt毎のユーザ集合および属性の変動の大きいDBが作ることができる．共有するDBを動的に変更し，かつ，仮想的なDBに対して$D_F$値を計算し提示する属性の組を決めれば，プロバイダに開示できる属性の組がΔt前後で変化するため，ユーザIDの特定を回避することができる．

### 5.2 関連研究

Borriellorら[5]はネットワーク上にあるprivacy proxyを介して，ユーザとプロバイダのエンドエンドにおける公開可能なプライバシ情報のネゴシエーションを行っている．本方式は，ユビキタスネットワー

キング環境でのS-privacyの使用 については述べているが，生体情報やユーザの行動も含めたD-privacyに対しては言及していない．D-privacyをプライバシ情報に含めた場合，本方式で実現した場合，情報量とプロバイダのサービス提供タイミングの増加に伴い公開情報ネゴシエーションの回数が増加するため，処理が煩雑になる．LooMでは，ここのユーザ／プロバイダ間ではネゴシエーションを行わず，簡単なエントロピー計算で情報提供可否を決定しているため，D-privacyが追加されても，利便性が低下することはない．

また，本方式はサービス起動に直接関係ない情報が入手可能なため，不正流出や別目的の使用の危険性がある．提案方式でも，悪意のあるプロバイダが入力インタフェース情報だと偽って，サービス起動とは直接関係のないプライバシ情報を入手可能である．プロバイダから入力インタフェース情報をサーバが取得し，必要な情報のみを返送する方式[6]などと組み合わせることで危険度が低下するが，その処理を行うためのprivacy proxyへ負荷が集中してしまう．LooMでは，プロバイダは，サービス起動に無関係なプライバシ情報も入手できるが，ユーザ個人に対しても，規定の匿名性を確保できるので，双方にとって利点がある．

ユーザIDとプライバシ情報を分離して，サービス起動制御を行う方式は，CONSORTS[3],[4]やCal-$(IT)^2$[2]でも行われている．ユーザIDを用いず，位置情報をトリガにしてサービスを提案／提供するものであり，研究の狙いは提案方式と似ている．しかし，位置情報以外のD-privacy 情報，およびプライバシ情報の組み合わせによるユーザIDの推測の観点からは，これらの研究は特に言及していない．

DBの中からあらかじめ関係属性に関連性を静的に調べ，属性マップを生成する．その属性マップが適当か否かをエントロピーで計算するという方式[9]がある．本方式は，DBの構成やユーザ数の変化を考慮していないため，本稿で前提としているような事例集合が時間とともに変化するような環境への応用は困難である．LooMは，マップを用いず，正規化した値だけを用いてプライバシ保護の評価基準を決定しているため，ユビキタスネットワーキング環境への適用が簡単なアルゴリズムで実現できる．

## 6. ま と め

ユビキタスネットワーキング環境におけるプライバシ保護とサービスの利便性を適度なバランスに保ちつつ，匿名性に対して定量的な評価尺度を持たせるためのLooMを提案し，シミュレーションによる評価，プロトタイプ実装によるフィージビリティの確認を行った．LooMは，ユビキタスネットワーキングにおけるプライバシ保護についてモデル化し，定量的評価を行った．エントロピーを匿名性の評価尺度に利用し，正規化によりユーザ集合の違いを隠蔽した．また，プロトタイプでは，属性数に応じて，エントロピー計算の方法を変えることで，適度なレスポンス性能を確保できることを示した．

今後は，正規化手法の詳細評価，　LooMサーバへの機能追加を行っていく．

## 文　献


[1] 今田美幸, “移動エージェントシステムにおけるセキュリティ実現方式,” 信学論, vol.J84-B, no.5, pp.932-939, 2001.

[2] CAL-$(IT)^2$, ”http://www.calit2.net/”.

[3] 辛島明男, 和泉憲明, 車谷浩一, 中島秀之, “ユビキタス環境におけるコンテンツ流通のためのマルチエージェントアーキテクチャ:CONSORTS,” 情処研報 Vol.2003, No.39, pp.7-14, 2003.

[4] 中島秀之, 橋田浩一, 森彰, 伊藤日出男, 本村陽一, 車谷浩一, 山本吉伸, 和泉潔, 野田五十樹, “情報インフラに基づくグラウンディングとその応用－サイバーアシストプロジェクトの概要－”, コンピュータソフトウェア, vol. 18, no.4, pp.48-56, 2001.

[5] G. Borriello, L.E. Holmquist, “ A Privacy Awareness Systemfor Ubiquitous Computing Environments,” In 4th International Conference on Ubiquitous Computinng (UbiComp2002), Springer-Verlag LNCS 2498, pp. 237-245, Sept. 2002.

[6] 田丸修平, 岩谷晶子, 高汐一紀, 徳田英幸, “プライバシを考慮した個人情報に適応的なアプリケーションフレームワーク,” 情処技報 vol. 2003, no.42, pp.49-56, 2003.

[7] J. R. Quinlan, J. Ross Quinlan, “C4.5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning),” Morgan Kaufmann Pub., 1993.

[8] MySQL, “http://www.mysql.com/”.

[9] LiWu Chang, Ira S. Moskowitz, “ An Integrated Framework for Database Privacy Protection ,” Data And Applications Security: Developments and Directions (eds. Thuraisingham, van de Riet, Dittrich and Tari), Kluwer Academic, pp 161-172, 2001.