

Web アクセス履歴における仮 ID の更新頻度に関する考察

福嶋 雄也^{1,a)} 北島 祥伍¹ 満保 雅浩²

概要: 改正個人情報保護法の施工に伴い、ビッグデータの積極的な利活用促進を目的として「匿名加工情報」制度が導入されたが、このような状況の中でプライバシー保護に関する法的・技術的検討はより重要になっていくと考えられる。履歴データに対する仮名化を考える際、長期間同一の仮 ID を使用することで再識別リスクが高まる可能性があることから、適切な頻度で仮 ID を更新するなどの処理を講じる必要がある。本論文では、履歴データの中でも特に Web ページのアクセス履歴について、集合間の類似度に基づき仮 ID の更新頻度の安全性について定量的な評価を行う。

キーワード: プライバシー保護, 仮名化, 安全性評価, Web アクセス履歴

A Study on Update Frequency of Pseudonym ID on Web Browsing History

KAZUYA FUKUSHIMA^{1,a)} SHOGO KITAJIMA¹ MASAHIRO MAMBO²

Abstract: In May 2017, the system "Anonymously Processed Information" was introduced to promote active utilization of big data and to ensure security and privacy of data. Under such circumstances, legal and technical consideration on privacy preserving will become more important. In case of considering pseudonymization of historical data, it is necessary to update the Pseudonym ID at an appropriate frequency, since using the same Pseudonym ID for a long time may increase the risk of re-identify. In this paper, we evaluate the relationship between update frequency of the Pseudonym ID quantitatively on the Web access history, based on the similarity between sets of the Web access data.

Keywords: Privacy Preserving, Pseudonymization, Safety Evaluation, Web-Browsing History

1. はじめに

1.1 背景

近年、企業や政府が保有するオープンデータ・ビッグデータの利活用が盛んに行われている。これらのデータにはパーソナルデータが含まれる場合もあり、個人のプライバシー保護に配慮した適切な取り扱いが求められる。平成 29 年 5 月には改正個人情報保護法が全面施工され、その中でビッグデータの積極的な利活用促進に寄与することを目的

とした「匿名加工情報」制度が導入された。このような状況の中、プライバシー保護に関する法的・技術的対策の検討はより重要性を増していくと考えられる。

データの匿名加工手法の一つとして仮名化が挙げられる。平成 28 年 8 月に経済産業省が公表した『匿名加工情報作成マニュアル』[1]においても、加工手法の一例として識別子の仮名化が示されている。一方で、仮名化自体は k -匿名化 [2] 等の匿名化手法と比べて非常に弱いものであり、また、仮名化に関する安全性の評価を行った研究として [3] 等があげられるが、これまで積極的に行われることはなかった。

1.2 本論文の目的

本論文では、Web アクセス履歴を対象とし、仮 ID の安

¹ 金沢大学大学院 自然科学研究科
Graduate School of Natural Science and Technology,
Kanazawa University

² 金沢大学 理工研究域
Institute of Science and Engineering, Kanazawa University

a) k-fukushima@stu.kanazawa-u.ac.jp

全性について更新頻度の観点から考察する。

具体的には、実際に収集した Web アクセス履歴に対して仮 ID の更新頻度を何種類か設定し、各仮 ID に対応するアクセス履歴間の類似度を利用した再識別手法を用いて仮 ID が再識別される割合を評価することで、設定した更新頻度の安全性について考察する。

2. 匿名加工と仮名化

データの匿名加工手法として k -匿名化 [2]、一般化や属性削除などの手法が挙げられるが、本論文ではそのうち仮名化に注目する。

2.1 仮名化

あるユーザの識別子を、容易に復元することが出来ない別の文字列に置き換える手法を仮名化といい、置換した文字列を仮 ID という。EU 一般データ保護規則第 4 条 5 号では、仮名化は以下のように定義されている。

(5) ‘pseudonymisation’ means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person;

また [1] では、仮名化は以下の性質を満たすべきだとされている。

不可逆性

仮 ID から、仮名化前の識別子が特定されないようにする必要がある。具体的には、ハッシュ化や UUID(Universally Unique Identifier) を用いる手法等が考えられる。

不連続性

履歴データや継続的にデータを提供する場合、同一の識別子に対して割り振られる仮 ID はある一定の期間で更新する必要がある。

非同一性

異なるデータ間において、同一の識別子に対して同一の仮 ID が割り振られないようにする必要がある。

また、同一の識別子に対して複数の仮 ID が割り振られるような仮名化を特に多重仮名化とよぶ。多重仮名化を行う際には、仮 ID の有効期間（更新頻度）を事前に設定し、その期間を超えた際には別の仮 ID を用いるという手法が一般的である。

Web アクセス履歴に対する仮名化及び多重仮名化の例を以下に示す。

表 1 アクセス履歴のサンプル

UserName	URL	AccessDate
Alice	www.google.co.jp	08/21 23:52:39
Bob	www.kanazawa-u.ac.jp	08/21 23:54:11
Alice	www.google.co.jp/maps	08/21 23:55:40
Carol	twitter.com/CSS2017Official	08/21 23:58:21
Bob	acanthus.cis.kanazawa-u.ac.jp	08/21 23:59:02
Alice	mail.google.com/mail	08/22 00:00:36
Carol	twitter.com/ipsjcom	08/22 00:01:10
Carol	www.facebook.com	08/22 00:03:56

表 2 仮名化を行ったデータ

PID	URL	AccessDate
00300	www.google.co.jp	08/21 23:52:39
00104	www.kanazawa-u.ac.jp	08/21 23:54:11
00300	www.google.co.jp/maps	08/21 23:55:40
00212	twitter.com/CSS2017Official	08/21 23:58:21
00104	acanthus.cis.kanazawa-u.ac.jp	08/21 23:59:02
00300	mail.google.com/mail	08/22 00:00:36
00212	twitter.com/ipsjcom	08/22 00:01:10
00212	www.facebook.com	08/22 00:03:56

表 1 は 3 人のユーザのアクセス履歴の例であり、表 2 は表 1 のユーザ名を仮名化したデータである。次に、仮 ID の更新頻度を時刻 00:00:00 を基準とし 24 時間に設定した多重仮名化データを表 3 に示す。

表 3 多重仮名化を行ったデータ

PID	URL	AccessDate
00300	www.google.co.jp	08/21 23:52:39
00104	www.kanazawa-u.ac.jp	08/21 23:54:11
00300	www.google.co.jp/maps	08/21 23:55:40
00212	twitter.com/CSS2017Official	08/21 23:58:21
00104	acanthus.cis.kanazawa-u.ac.jp	08/21 23:59:02
00152	mail.google.com/mail	08/22 00:00:36
00435	twitter.com/ipsjcom	08/22 00:01:10
00435	www.facebook.com	08/22 00:03:56

表 2 では、Alice : 00300, Bob : 00104, Carol : 00212 と仮名化されているのに対し、表 3 では Alice : {00300, 00152}, Bob : 00104, Carol : {00212, 00435} と多重仮名化されており、データ上 Alice 及び Carol のアクセス履歴は 2 人のユーザに分割されている。また、本論文では一つの仮 ID に対応する識別子は一つのみとし、複数の識別子が同時に一つの仮 ID に割り当てられるような場合は考慮しない。

2.2 仮名化の利点及び問題点

2.1 節で例示したように、多重仮名化を行うことで、同一ユーザの履歴もデータ上は別のユーザの履歴ということになっているため、個人識別に係るリスクを低減することができる。一方、更新頻度を大きくすると（すなわち仮 ID の有効期間を短くとると）データの解析を行う際の有用性が低下すると考えられ、この意味において安全性と有用性はトレードオフの関係にある。

平成 25 年、JR 東日本が Suica の乗降履歴データを外部販売しようとした一件では、同社は仮 ID の更新期間を 1 か月に設定していたと発表したが、この更新期間の設定は安全でないのではないかとという批判が多く寄せられた。この件からもわかるように、データの特徴に応じた仮 ID の更新頻度の定量的な評価は重要な課題であるといえる。

3. 仮 ID 更新頻度の評価手法

本論文では、PWS CUP 2016[4] で多くのチームが用いた Jaccard 係数による再識別手法を参考とし、仮 ID の更新頻度の評価に用いる。PWS CUP では、匿名加工された顧客データと元の顧客データの対応関係を推測する際、各顧客データに対応するトランザクションデータ間の類似度を評価して再識別を行ったが、本論文では、各仮 ID に対応する Web アクセス履歴間の類似度を用い、多重仮名化されたデータにおいて、ある仮 ID と同一のユーザに割り当てられた他の仮 ID を推測する。これを事前に設定するいくつかの更新頻度に対して行い、ある更新頻度において、同一ユーザに割り当てられた仮 ID が平均してどの程度推測されるかを評価する。

3.1 データ構造の定義

本論文で用いるアクセス履歴データ D の構造を次のように定義する： D は N 行からなり、そのうち i 行目は

$$D_i = (ID_i, URL_i, Date_i)$$

の形で書けるとする。ここで $ID_i \in \mathbf{ID}$, $URL_i \in \mathbf{URL}$, $Date_i$ はそれぞれユーザの識別子、アクセスした URL 及びアクセスした日時を表す。また、ある識別子 ID_i を持つユーザが D 中でアクセスした URL の全体集合を $\mathbf{URL}(ID_i)$ と表す。

3.2 仮名化・多重仮名化の定義

ある $ID_i \in \mathbf{ID}$ と仮名化アルゴリズム P に対し、 $P(ID_i) = PID_i \in \mathbf{PID}$ を仮名化という。ここで \mathbf{PID} は仮 ID の全体集合を表す。また、 ID_i に割り当てられる仮 ID が $S(i)$ 個存在するとき、すなわち $PID_i = \{PID_i(1), \dots, PID_i(S(i))\} \subseteq \mathbf{PID}$ となる場合を特に多重仮名化という。ここで、 $S(i)$ は ID_i に対応する仮 ID の

割り当て数を表す。

3.3 仮 ID の再識別と更新頻度の評価

次に示すアルゴリズム 1 では、Jaccard 係数を用いてある 2 つの仮 ID に対応するアクセス履歴間の類似度を測定し、その値が大きいものを候補として選択する手法で再識別を行っている。

また、ある二つの仮 ID x, y が同一のユーザに割り当てられているとき 1、そうでない時 0 を取る関数 $f(x, y)$ があるとする。ここでは、評価の都合上 f として $x \equiv y \pmod{n}$ (n : 全ユーザ数) ならば $f(x, y) = 1$ となるように仮 ID を割り当てるものとする。また、ある仮 ID x と同一のユーザに割り当てられている仮 ID の数 (x 自身を除く) を $s(x)$ と表す。

Algorithm 1 多重仮名化の Jaccard 再識別

Require: $\mathbf{PID}, \mathbf{URL}(\mathbf{PID})$, ユーザ数 n

Ensure: 再識別率 $Reid$

Step 1.

$\forall x, y \in \mathbf{PID} (x \neq y)$ に対し Jaccard 係数

$$J(x, y) = \frac{\mathbf{URL}(x) \cap \mathbf{URL}(y)}{\mathbf{URL}(x) \cup \mathbf{URL}(y)}$$

を計算する。

Step 2.

$\forall x \in \mathbf{PID}, \forall y \in \mathbf{PID}/\{x\}$ に対し

$C_x \leftarrow \{y \mid \text{上位 } s(x) \text{ 個の } J(x, y) \text{ を取る } y\}$
とする。

Step 3.

$\forall x \in \mathbf{PID}, \forall y \in C_x$ に対し

$$Reid_x = \frac{|\{y \mid f(x, y) = 1\}|}{|C_x|} = \frac{|x \equiv y \pmod{n}|}{|C_x|}$$

アルゴリズム 1 を用いてすべての仮 ID に対する再識別率 $Reid$ を求めることができる。この結果を用い、ある更新頻度における平均再識別率を次のように定義する。

定義 1 (平均再識別率) ある更新頻度に対して仮 ID の全体集合を \mathbf{PID} としたとき、平均再識別率を

$$\frac{1}{|\mathbf{PID}|} \sum_{x \in \mathbf{PID}} Reid_x = \frac{1}{|\mathbf{PID}|} \sum_{x \in \mathbf{PID}} \frac{|\{y \mid f(x, y) = 1\}|}{|C_x|}$$

と定める。

ここで、アルゴリズム 1 の動作例を示す： $\mathbf{PID} = \{a, b, c, d, e, f, g\}$ であるとし、仮 ID a と同一のユーザに割り当てられる他の仮 ID の候補を選択し再識別率を評価する。ここで、 $s(a) = 3$ 、すなわち a の他に 3 つ、同一のユーザに対応する仮 ID が存在しているとする。

Step 1 では a と a 以外すべての仮 ID との Jaccard 係数を求める。ここでの計算結果は次に示す表 4 のとおりであ

表 4 Jaccard 係数の計算結果例

PID	Jaccard
b	0.7
c	0.8
d	0.5
e	0.75
f	0.4
g	0.25

るとする。

Step 2 では、 a との Jaccard 係数が大きい上位 $s(a) = 3$ 個の仮 ID を候補 C_a として選択する。ここでは、表 4 より $C_a = \{b, c, e\}$ となる。

Step 3 では再識別率を評価する。ここで $C_a = \{b, c, e\}$ に対し $a \equiv b \equiv e \pmod{n}$ であるとする、 a の再識別率は $Reid_a = 2/3$ となる。

4. シミュレーション手法と結果及び評価

4.1 シミュレーション手法

103 人のユーザから収集された、2016/12/10 から 12/16 の 1 週間分のアクセス履歴を分析に用いる。収集されたデータの各レコードは 3 つの要素 $ID, URL, Date$ からなり、それぞれユーザの識別子、アクセスした URL、アクセスした日時 (JST) を表す。なお URL について、今回のシミュレーションではフルパスの URL とドメイン (FQDN) の 2 通りを対象とする。また、仮 ID の更新頻度は 24, 12, 8, 6, 4, 3, 2, 1 時間の 8 通りとし、事前に設定したある更新頻度に基づき、時刻 00:00:00 を基準として仮 ID を更新することとする。

まず、設定したある更新頻度に対して生成された仮 ID の相異なる二つの組合せ全通りに対して Jaccard 係数を求めることで、ある更新頻度においてユーザ間のアクセス履歴の類似度がどの程度になるかを評価する。

次に、アルゴリズム 1 で示した再識別手法を用い、すべての仮 ID について、それと同一のユーザに割り当てられる仮 ID の候補を選択し、実際に正しい仮 ID が選択されている割合を求めることで、ある更新頻度についてどの程度の割合で仮 ID の再識別がなされるかを評価する。

4.2 シミュレーション結果

4.2.1 アクセス履歴の類似度

まず、ドメインを対象とした解析の結果を示す。表 5 は、アルゴリズム 1 の Step 1 において導出した Jaccard 係数の平均値を示したものであり、ここで「全通りの組合せ」とは相異なる二つの仮 ID の組全通りの Jaccard 係数の平均値であり、「同一ユーザ同士の組み合わせ」とは同一ユーザに対応する仮 ID 同士の組全通りの Jaccard 係数の平均値である。また図 1, 2 はそれぞれ更新頻度 24 時間または

表 5 Jaccard 係数の平均値 (ドメイン)

更新頻度 (時間)	全通りの組合せ	同一ユーザ同士の組合せ
24	0.1252	0.3528
12	0.1220	0.3216
8	0.1181	0.3041
6	0.1139	0.2927
4	0.1114	0.2890
3	0.1067	0.2782
2	0.1032	0.2735
1	0.0923	0.2662

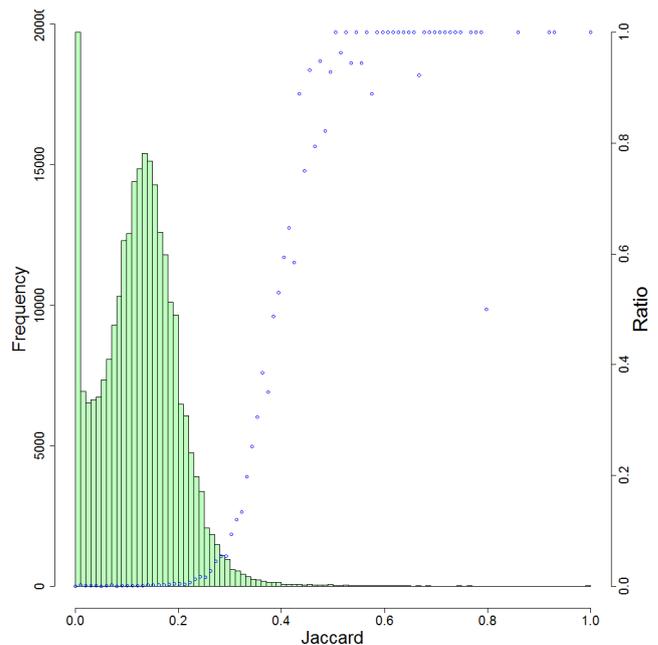


図 1 Jaccard 係数の度数分布 (ドメイン, 更新頻度 24 時間)

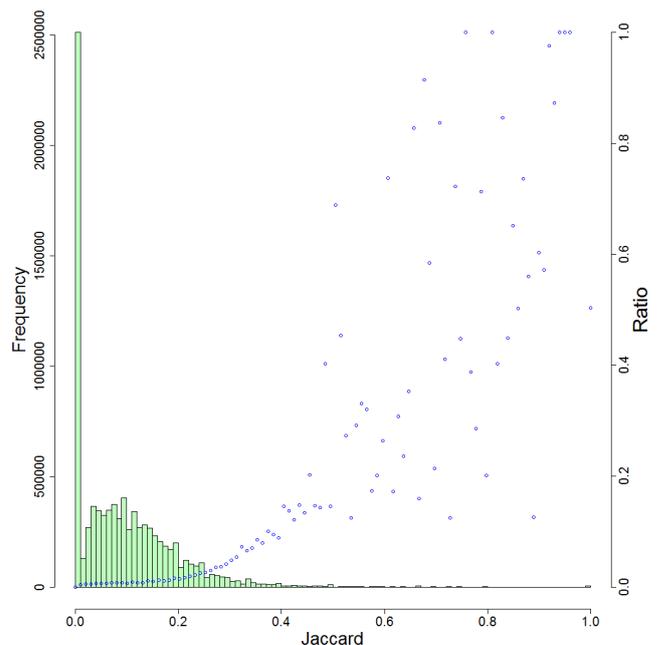


図 2 Jaccard 係数の度数分布 (ドメイン, 更新頻度 1 時間)

1 時間の際の「全通りの組合せ」の Jaccard 係数の度数分布を示しており、グラフ上の点は各度数中の「同一ユーザ同士の組合せ」の割合を表している。

今回設定した 8 通りの更新頻度に対して、「全通りの組合せ」、「同一ユーザ同士の組合せ」いずれも更新頻度が高くなるほど平均値が減少しているが、「同一ユーザ同士の組合せ」の方がより大きく値が減少している。また、「全通りの組合せ」における Jaccard 係数の平均値の減少は、更新頻度が高いほど Jaccard 係数の小さい組合せの割合が増加しているためとも考えられる (図 1, 2)。これらの結果は、更新頻度が低いほどアクセス履歴の類似性が平均的に大きくなっていることを示している。また図 1, 2 より、更新頻度が 24 時間の場合では Jaccard 係数が 0.5 以上の組合せはそのほとんどが「同一ユーザ同士の組合せ」であるのに対し、更新頻度 1 時間の場合では Jaccard 係数が 0.5 以上の組合せでも「同意ユーザ同士の組合せ」はそれほど支配的でなく、平均して半分程度の割合となっている。

次に、フルパスの URL を対象とした解析の結果を示す。

表 6 Jaccard 係数の平均値 (フルパス)

更新頻度 (時間)	全通りの組合せ	同一ユーザ同士の組合せ
24	0.00581	0.0422
12	0.00641	0.0411
8	0.00646	0.0408
6	0.00634	0.0405
4	0.00646	0.0414
3	0.00642	0.0409
2	0.00634	0.0403
1	0.00610	0.0419

表 6 は各更新頻度における Jaccard 係数の平均値を、図 3, 4 はそれぞれ更新頻度 24 時間と 1 時間の場合における Jaccard 係数の度数分布を表している。

フルパスの場合、「全通りの組合せ」、「同一ユーザ同士の組合せ」いずれにおいても更新頻度に関する相関はほぼ見られず、また特に「全通りの組合せ」において Jaccard 係数の平均値が非常に低くなっている。これは、異なる仮 ID に対応するアクセス履歴間で URL が完全一致する例が極端に少ないためであると考えられる。一方で、更新頻度 24 時間の場合では Jaccard 係数が 0.1 以上、1 時間の場合では 0.4 以上の組合せはそのほとんどが「同一ユーザ同士の組合せ」となっており、ドメインを対象とした場合と比べ全体としては Jaccard 係数の平均値は低いものの、ユーザのアクセス履歴の類似性はより大きく表れている。

次に示す図 5, 6 はそれぞれ「全通りの組合せ」、「同一ユーザ同士の組合せ」におけるフルパス URL とドメインについての更新頻度ごとの Jaccard 係数の変化を表したグラフである。

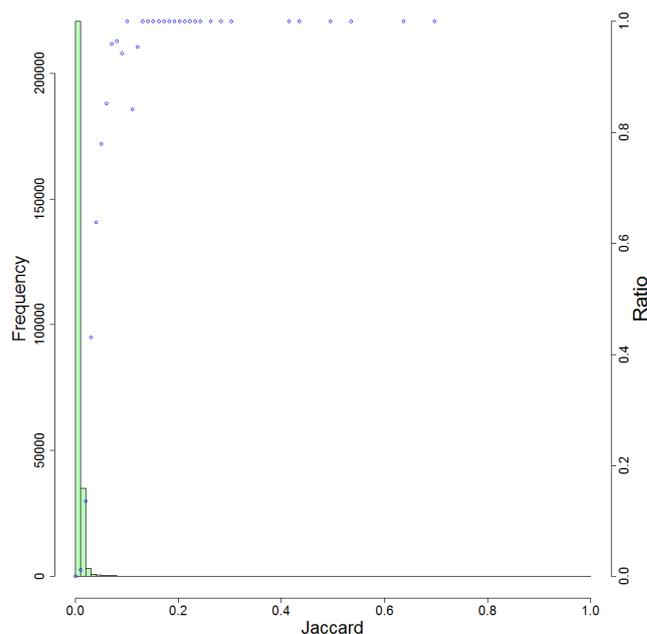


図 3 Jaccard 係数の度数分布 (フルパス, 更新頻度 24 時間)

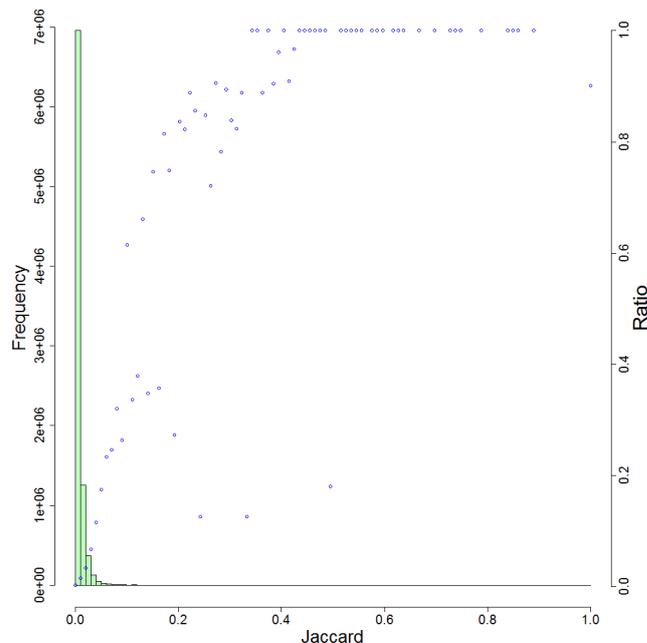


図 4 Jaccard 係数の度数分布 (フルパス, 更新頻度 1 時間)

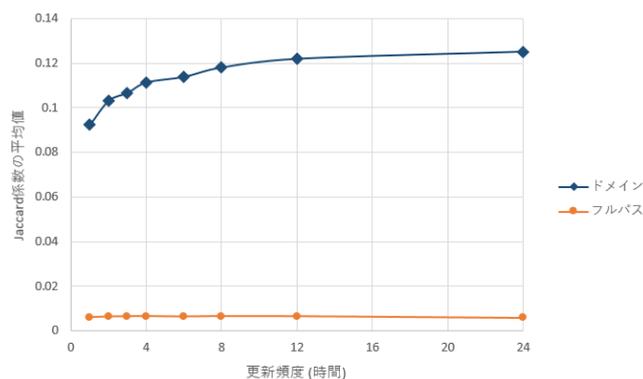


図 5 「全通りの組合せ」における Jaccard 係数の平均の変化

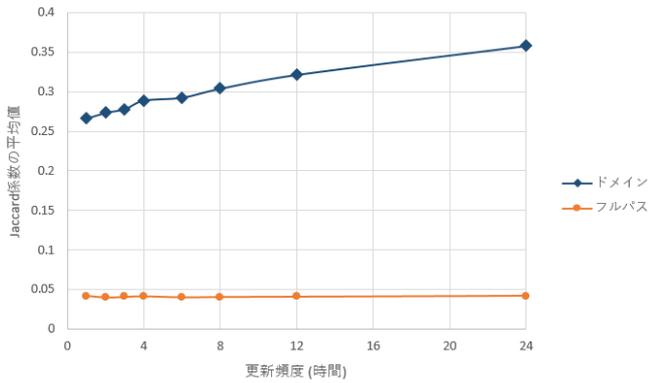


図 6 「同一ユーザ同士の組合せ」における Jaccard 係数の平均の変化

4.2.2 平均再識別率

表 7 平均再識別率

更新頻度 (時間)	平均再識別率 (ドメイン)	平均再識別率 (フルパス)
24	0.6309	0.7063
12	0.4817	0.5544
8	0.4294	0.4971
6	0.3896	0.4608
4	0.3630	0.4230
3	0.3341	0.3889
2	0.3083	0.3585
1	0.2688	0.2950

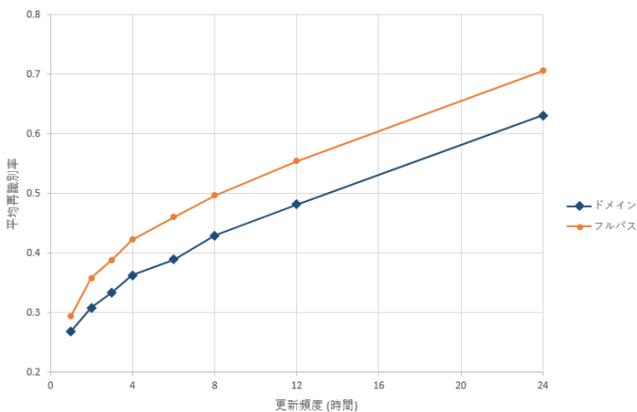


図 7 平均再識別率の変化

各更新頻度に対するドメイン、フルパスそれぞれを対象とした場合の平均再識別率を表 7 に示す。あわせて、平均再識別率の変化を表すグラフを図 7 に示す。これらの図表から明らかなように、ドメイン、フルパスいずれの場合でも更新頻度が大きくなるほど平均再識別率は徐々に低下している。また、どの更新頻度においてもフルパスを対象と

した場合の方がドメインを対象とした場合よりも高い平均再識別率を示している。

ここで、更新頻度が 24 時間の場合の再識別率の度数分布をドメインを対象とした場合とフルパスを対象とした場合についてそれぞれ図 8, 9 に示す。同様に更新頻度 1 時間の場合の再識別率の度数分布についてもドメインとフルパスを対象とした場合をそれぞれ図 10, 11 に示す。

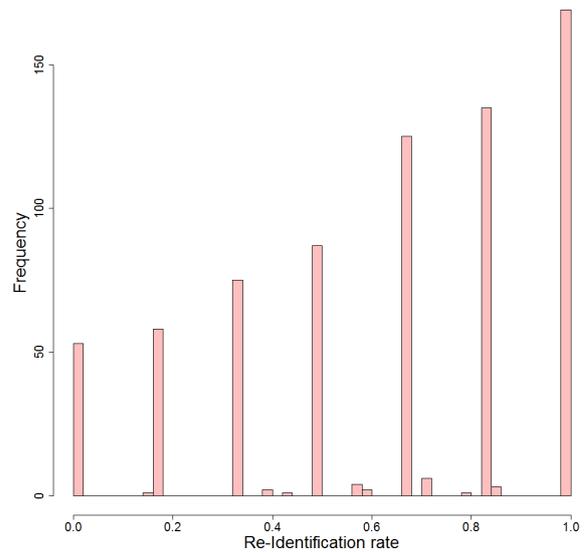


図 8 再識別率の度数分布 (ドメイン, 更新頻度 24 時間)

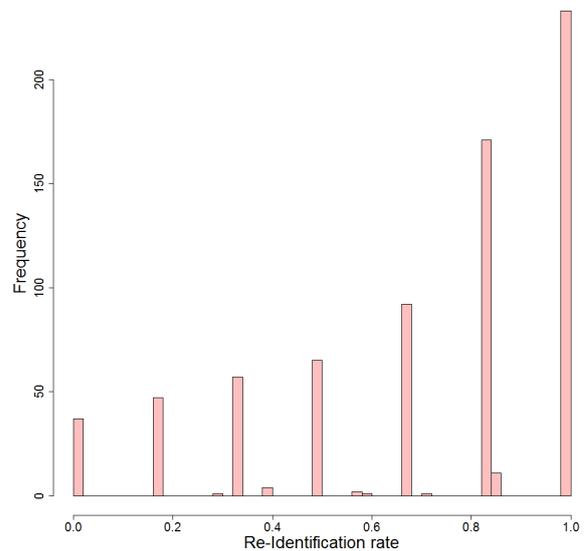


図 9 再識別率の度数分布 (フルパス, 更新頻度 24 時間)

更新頻度 24 時間の場合においては全 720 個の仮 ID に対して $Reid = 1$ となる仮 ID が 178 個存在し、全体のおよそ 25% の仮 ID について同一ユーザに対応する他の 6 個の仮 ID がすべて再識別される結果となった。また、更新頻度

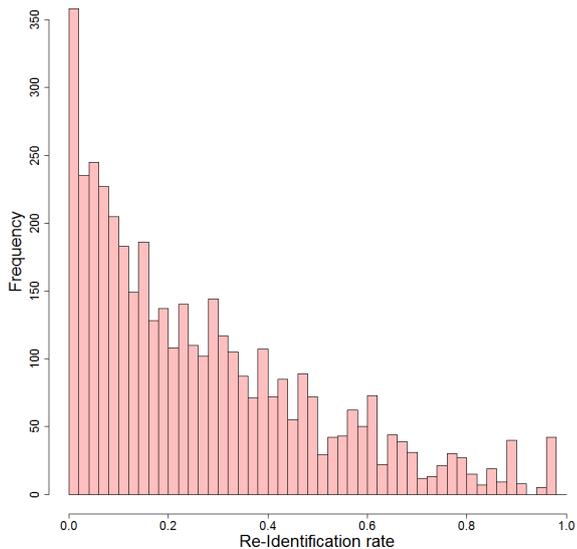


図 10 再識別率の度数分布 (ドメイン, 更新頻度 1 時間)

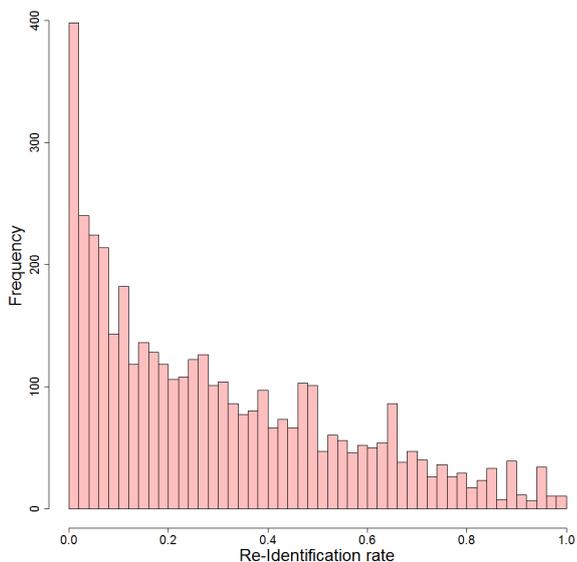


図 11 再識別率の度数分布 (フルパス, 更新頻度 1 時間)

が 1 時間の場合でも平均再識別率は 0.2976 であり, 平均して 3 割程度の仮 ID が再識別されることがわかる。

また, 図 8, 9 及び図 10, 11 の比較からフルパスの URL を解析対象とした場合でも再識別率の分布はほぼ同様の割合になっているが, 更新頻度 24 時間の場合では平均して 7 割程度の仮 ID が再識別されるためよりリスクが大きいといえる。

4.3 分析結果の考察

4.3.1 対象とするアクセス履歴の種類

本研究では, アクセス履歴データとしてフルパスの URL 及びそのうちドメイン (FQDN) 部を抜き出したものの 2 種類について分析を行った。4.2.1 では Jaccard 係数を用いて

アクセス履歴間の類似度を求めており, 図 5, 6 から分かるようにフルパス URL を対象とした場合は「全通りの組合せ」, 「同一ユーザ同士の組合せ」いずれの場合においてもドメインを対象とした場合を下回っているが, 一方図 7 から分かるようにどの更新頻度に対してもフルパス URL を対象とした場合の方が高い平均再識別率を示している。このように「同一ユーザ同士の組合せ」の Jaccard 係数の平均値が低い場合でも再識別の精度が高くなる場合もあり, 単純な類似度による評価のみでなく, 今回のように何らかの再識別手法等を利用した評価も組み合わせることは有用と言えよう。

4.3.2 アクセス履歴の類似性と匿名性

表 5, 6 では, Jaccard 係数を用いたアクセス履歴間の類似度を求めている。今回設定した 8 通りの更新頻度いずれに対しても「全通りの組合せ」と「同一ユーザ同士の組合せ」では分布に大きな差が見られ, Web のアクセス履歴についてユーザ毎に閲覧するサイトの特徴が大きく出ていることが分かる。よって, 仮 ID の更新頻度を適切に定めないと, 同一ユーザに対応する仮 ID の組を容易に再識別されかねないが, 本論文の結果から, 更新頻度が 24 時間未満の場合において更新頻度を大きくすることで「同一ユーザ同士の組合せ」について平均再識別率は低下させることができ, 匿名性は向上していると考えられる (図 5)。

一方, 更新頻度を大きくとることで一つの仮 ID に対応するアクセス URL の集合のサイズが極端に小さくなる場合があり, これはデータの有用性の低下に結びつきうる。

5. まとめと今後の課題

本論文では, Jaccard 係数を用いた再識別手法によりある更新頻度に基づいて仮 ID の更新を行った際, 同一ユーザに対応する仮 ID がどの程度再識別されるか評価するアルゴリズムを提案し, 実データに対して適用することでその有用性の評価を行った。その結果, ドメインとフルパス URL を解析対象とした場合ではフルパス URLの方がより再識別精度が高く, 更新頻度が 24 時間の場合では平均して 7 割程度が再識別されることを示した。あわせて, アクセス履歴間の類似度の平均値が低い場合でも, 平均再識別率自体は大きくなるような例を示し, 単純な類似度のみでなく再識別手法等を用いた評価の併用も有用であることを示した。

今後の課題として, Jaccard 係数以外の類似度評価指標の適用や異なる仮 ID の割り当て方の検討, また実際に指標をデータに適用する際, 平均再識別率の閾値をどのように定めるかという点等が挙げられる。

謝辞 本研究成果の一部は, 国立研究開発法人情報通信

研究機構 (NICT) の委託研究「Web 媒介型攻撃対策技術の実用化に向けた研究開発」により得られたものです。

参考文献

- [1] 経済産業省, ” 事業者が匿名加工情報の具体的な作成方法を検討するにあたっての参考資料 (「匿名加工情報作成マニュアル」) Ver1.0”, 2016 年 8 月. [<http://www.meti.go.jp/press/2016/08/20160808002/20160808002-1.pdf>, 2017 年 8 月 25 日参照]
- [2] L. Sweeney, ”k-anonymity: a model for protecting privacy”, *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5), pp. 555-570, 2002.
- [3] 早稲田 篤志, 野島 良, 盛合 志帆, 菊池 浩明. ” 良い仮名化悪い仮名化”, 2017 年暗号と情報セキュリティシンポジウム (SCIS2017), 2D1-6, 2017
- [4] 菊池 浩明, 小栗 秀暢, 野島 良ら. ”PWSCUP: 履歴データを安全に匿名加工せよ”, *コンピュータセキュリティシンポジウム 2016 論文集*, pp.271-278, 2016.