ユーザの日本語ローマ字入力特性を用いたシグネチャ型なりすまし検知システムの提案

関口 博幸† 勅使河原 可海† 西垣 正勝‡
†創価大学大学院工学研究科 ‡静岡大学情報学部

1. はじめに

近年、インターネットが世界的規模で飛躍的な発展を遂げてきている。それに伴って、企業などにおける基幹ネットワークが、インターネットに直結されるようになった。その一方で、コンピュータ端末やネットワークのログインにおいて、パスワードが容易に破られてしまう事件が多発しているのも事実である。このような現状において、他人受入率の低減を目的としたバイオメトリクスなどの個人認証技術が研究されている。しかしながら、グミフィンガーに代表されるように、個人の特徴を利用した個人認証技術を、回避するための技術も研究されてきている[1]。このように、現在の個人認証技術の多くは、他人受入率の低減を目的としたものであり、個人認証後のことは考慮されていない。したがって、個人認証を回避された場合には、そのシステムが極めて無防備な状態になることから、個人認証後における対処が求められる。

これまで、ユーザのコンピュータ利用に現れる習慣や癖などの特性を用いて、なりすましを検知する方法を検討してきた。本研究では、複数の利用特性を要素として用いることにより、誤検知の低減を図ることを目的としている。本稿では、認証後において、要素のひとつとして、日本語ローマ字表記の差異をユーザの入力特性として用い、その入力特性を正規ユーザのシグネチャと比較することにより、なりすましを検知するシステムを提案する。

2. 既存のなりすまし検知技術の問題点

前述したように、個人認証を回避された場合には、そのシステムが極めて無防備な状態になることから、個人認証後における対処が求められる。そこで考えられた技術が、なりすまし検知技術である。しかしながら、既存のなりすまし検知技術には、以下のような問題点がある。

1) 外部犯を対象としている

既存のなりすまし検知技術は、侵入検知技術の延長であり、主に外部犯を対象としている。また、現在のOSの多くがGUI環境を基本としているにもかかわらず、GUI 環境を考慮しているものは少ない。したがって、内部犯に対する対処が不完全である。

2) 検出に曖昧さが残る

既存のなりすまし検知技術は、ユーザの正常状態をプロファイルとして定義し、その定義からの逸脱を検出する異常検出の手法を用いている。また、一般的にプロファイルの準備には期間が必要である。さらに、プロファイルを用いた検出には幅があるため、常に曖昧さが残存する。

3. シグネチャ型なりすまし検出

検出手法は、不正な状態を定義したものであるシグネチャを用いる不正検出と、プロファイルを用いる異常検出に大きく分けられてきた。しかしながら、NPO 日本ネットワークセキュリティ協会（JNSA）が新たな分類を提案している[2]。表1に示すように、シグネチャを厳密なものとし、プロファイルを幅があるものとして分類している。また、検出手法のそれぞれに対して、シグネチャとプロファイルを用いる場合があるとしている。

表1 検出手法の分類

	Misuse Detection **不正検出**	Anomaly Detection **異常検出**
Signature based **(厳密)**	**不正**Signature**と一致** **→不正と判断**	**正常**Signature**と不一致** **→異常と判断**
Profile based **(幅あり)**	**不正**Profile**に近い** **→不正と判断**	**正常**Profile**から遠い** **→異常と判断**

提案システムのシグネチャ型なりすまし検知とは、このJNSAが提案する分類にのっとったものである。

4. 提案システム

提案システムでは、ユーザがローマ字入力モードで日本語を入力する際に、ユーザによって使用する日本語ローマ字表記に、差異が存在することに注目している。日本語ローマ表記には、表2に示した「訓令式」、「ヘボン式」や「日本式」などの表記法が存在し、これらの表記法を混合して使用しているユーザが多い。また、これらの表記法には存在しないが、標準的に採用されている表記法も使用されている。

提案システムでは、これらの日本語ローマ字表記

A Proposal of a Signature Base Spoofing Detection System Using Roman Inputting Patterns.
Hiroyuki Sekiguchi†, Yoshimi Teshigawara†, Masakatsu Nishigaki‡
†Graduate School of Engineering, Soka University
‡Faculty of Information, Shizuoka University

の中から、23 種類（53 項目）を選択して、ユーザ毎にそれぞれの日本語ローマ字表記について、使用するか否かを不正シグネチャとして定義している。この不正シグネチャをもとに、不正検出を行うことにより、厳密な検出を可能にしている。

表 2　訓令式

ア行	a	i	…		
カ行	ka	ki	…	kyu	kyo
サ行	sa	si	…	syu	syo
⋮	⋮	⋮	⋮	⋮	⋮
ガ行	ga	gi	…	gyu	gyo
ザ行	za	zi	…	zyu	zyo
ダ行	da	zi	…	zyu	zyo
バ行	ba	bi	…	byu	byo
パ行	pa	pi	…	pyu	pyo

5. 実験と評価

提案システムの有効性を検証するために、9 名のコンピュータユーザを対象に、検証プログラムをバックグラウンドで実行した状態で、ワープロソフトやメールソフトなどに対して入力を行ってもらった。全ユーザはコンピュータ利用経験が最低でも 3 年以上であり、常日頃からプログラム開発や研究のためにコンピュータを多用している大学生や大学院生である。

ユーザ毎の不正シグネチャを作成するために、全ユーザにアンケートを実施した。アンケートでは、使用するか否かと、同じ日本語に対して複数の表記を使用する場合を考慮して、その割合をユーザの判断で返答してもらった。このアンケート結果をもとに、利用割合が 5 割以上の場合に「使用する」とシグネチャを定義し検証を行った。また、実験で入力してもらった文章は､全ユーザで同一のものである。

表 3　検出結果（5 割以上）

	A	B	C	D	E	F	G	H	I
A	1	6	12	27	19	14	15	1	4
B	6	0	9	29	20	19	15	6	31
C	12	15	12	20	27	25	25	12	14
D	22	25	27	2	17	11	14	19	23
E	20	20	16	23	0	9	6	16	16
F	13	18	24	15	9	2	5	9	14
G	14	14	15	17	2	4	1	11	10
H	2	4	10	28	17	15	13	2	2
I	3	3	9	29	17	16	12	3	1

利用割合が 5 割以上の場合の検出結果を表 3 に示す。左端の列はどのユーザのシグネチャを使用したかを示し、上端の行はそのシグネチャに対するユーザを示す。5 割以上の場合では、ユーザ C のシグネチャに対する、本人の検出数と他人の検出数が同じ値になってしまい、本人と他人を区別することができなかった。この原因としては、使用割合をユーザの判断で返答してもらったが、本人の判断以上に使用していた表記があったためだと考えられる。そこで、同じ収集データを利用して 4 割以上の場合にも実験を行った。

表 4　検出結果（4 割以上）

	A	B	C	D	E	F	G	H	I
A	1	6	12	27	19	14	15	1	4
B	6	0	9	29	20	19	15	6	4
C	11	11	8	19	23	24	21	11	11
D	22	25	27	2	17	11	14	19	23
E	20	20	16	23	0	9	6	16	16
F	13	18	24	15	9	2	5	9	14
G	14	14	15	17	2	4	1	11	10
H	2	4	10	28	17	15	13	2	2
I	3	3	9	29	17	16	12	3	1

利用割合が 4 割以上の場合の検出結果を表 4 に示す。4 割以上の場合では、ユーザ C のシグネチャに対する本人の検出数が他人の検出数を下回り、本人と他人を区別することができた。しかしながら、ユーザ A と H のシグネチャを使用した場合では、低い検出数において同じ値が存在するため、本人と他人の区別に、さらに多くの入力が必要になると考えられる。

6. 今後の課題

今後、本研究を進めて行くには以下のような課題が残っている。

- シグネチャをユーザの判断ではなく、自動的に定義する方法について、さらに検討を進めて行きたい。
- ユーザ毎に不正と判断する検出数の閾値を、どのように決定するかについて、さらに検討を進めて行きたい。
- 有効性を示すために重要となる、本人と他人を区別するために必要な入力文章の平均長を、どのようにして求めるかついて、さらに検討を進めて行きたい。
- 日本語ローマ字入力特性が酷似するユーザが現れることを考えられるため、その他の利用特性について、さらに検討を進めて行きたい。

参考文献

[1] 山田浩二，松本弘之，松本勉：指紋照合装置は人工指を受け入れるか，ISEC2000，2000.7
[2] NPO 日本ネットワークセキュリティ協会 http://www.jnsa.org/