CL-003

キーストローク認証における k-means を用いたウルフ攻撃
Wolf Attack Using k-means in Keyboard Dynamics

泉 和哉 † 吉武 大地 ‡ 佐村 敏治 †
Kazuya Izumi Daichi Yoshitake Toshiharu Samura

1. はじめに

機密情報を扱う処理もネットワーク上で行われるようになり，ネットワーク上で飛び交う情報の重要性が高くなってきている．そのため，情報を送受信するユーザ自身が，本人であるかどうかの確認手段の 1 つとして，本人の生体的特徴を用いた生体認証が注目されており，実用化され普及してきている．

本研究では，生体認証の一種である非定型文キーストローク認証に着目する．キーストローク認証は，行動的特徴であるキーボードからの入力パターンを利用した認証であり，キーボードの入力パターンは各個人で異なるということを活用し，認証を行う．非定型文キーストローク認証の応用例としては，e-Learning システムなどが挙げられる．

生体認証が普及してきている一方で，生体認証特有の脆弱性が存在する．第三者によるなりすましは，生体認証に対する脅威のひとつである．生体認証において，複数の登録データに本人としてなりすませる人工的な入力データをウルフといい，ウルフを用いたなりすまし攻撃をウルフ攻撃と呼ぶ．

ウルフ攻撃は多くの登録データを少ない入力データでなりすましを行えるため，多くの正規ユーザが損害を被ることになる．また，一般的な従来の生体認証システムはウルフのような人工的なデータの入力を考慮していない．

ウルフ攻撃に関する多くの研究は，指紋認証や虹彩認証である[1]．指紋認証では，探索的に最も多くの登録ユーザになりすましに成功したデータをウルフとする手法や，各特徴量を多くの人間が持つ値に設定してウルフを合成する手法が提案されている．また，虹彩認証では，認証に用いられる虹彩の領域が極端に小さくなる特殊な画像を用いる手法が提案されている．この様に生体認証ごとにウルフを生成する様々な手法が提案されている．

キーストローク認証が実用化され，重要性は近年高まっており，多くの研究がなされている[2, 3]．しかしながら，キーストローク認証におけるウルフ攻撃に関する研究は我々の知るところ未だない．

本研究では，非定型文キーストローク認証におけるウルフ攻撃の手法について検討する[4]．キーストローク認証におけるウルフ攻撃のリスクを実証実験で明らかにし，今後キーストローク認証の安全性を向上させるための判断材料の 1 つとすることを本研究の目的とする．また，本研究のようなアプローチはキーストローク認証本来の強みや短所についての本質を追求することにもなる．

† 明石工業高等専門学校 NIT, Akashi College
‡ 京都大学 Kyoto University

2. キーストローク認証

キーストローク認証とは，ユーザのキーボードの入力パターンを生体情報として用いる行動的生体認証のことである．キーストローク認証は入力に用いる文字列の情報や，文字列の長さによって 3 種類の方式に分けられる．パスワードと組み合わせて，パスワードの入力パターンを用いた定型語キーストローク認証，特定の文字列の入力パターンを用いた定型文キーストローク認証，もう 1 つは入力される文字列に依存しない非定型文キーストローク認証である．非定型文キーストローク認証は入力文字数が多いため，比較的認証率が高い．そのため，本研究では後者の非定型文キーストローク認証を取り扱う．

非定型文キーストローク認証では，ユーザがシステムを利用する時のキーボードの入力パターンを用いて認証を行う．ユーザによる特別な操作なしで，システム利用中の継続的な認証が可能となる．そのため，e-Learning システムやメール，文書作成などで利用できる．

図 1 にキーストローク認証に用いる時間を示す．press はキーの押下時刻，release はキーの離上時刻を表す．そして，*pr* は 2 つのキーの押離時間，*pp* は 2 つのキーの押押時間，*rr* は 2 つのキーの離離時間，*rp* は 2 つのキーの離押時間である．

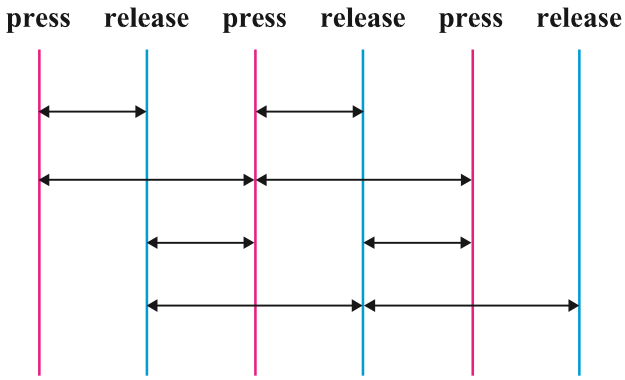

図 1 キーストローク認証に用いる時間

pr，*pp*，*rr*，*rp* それぞれについて，中央値の 2 倍以上の値と中央値の 0.5 倍以下の値を外れ値として削除し，各時間の平均値と標準偏差，最大値，最小値を求め，それぞれ 4 種類の特徴量抽出を行う．合計 16 個の特徴量を抽出し，認証に用いる．なお，特徴量の時間分解能をミリ秒とする．

本認証システムの識別関数には，ユークリッド距離とマンハッタン距離を用いる．キーストローク認証におけるウルフ攻撃に関する研究は，我々の知るところ初めての研究なので，識別関数として基本的であるそれら距離を用いてウルフ攻撃の被害にどのような差が出るのか比較を行う．

以降では，ユークリッド距離を用いた識別関数を *ED*, マンハッタン距離を用いた識別関数を *MD* と定義する．入力

Copyright © 2019 by Information Processing Society of Japan and
The Institute of Electronics, Information and Communication Engineers
All rights reserved.

データと登録データは，特徴量である *pr*，*pp*，*rr*，*rp* それぞれの平均値と標準偏差，最大値，最小値を要素として構成される 16 次元ベクトルとなる．閾値を *T* とし，入力データのベクトル要素を p_i，登録データのベクトル要素を q_i とする．識別関数が *ED* のとき式(1)，*MD* のとき式 (2) の条件を満たす場合に入力データを本人と見做し，そうでない場合は他人と見做す．

$$\sqrt{\sum_{i=1}^{16}(p_i - q_i)^2} \leq T \tag{1}$$

$$\sum_{i=1}^{16}|p_i - q_i| \leq T \tag{2}$$

3. ウルフ攻撃

ウルフとは，1 つの入力データで複数の登録データになりすますことができるデータのことである．このウルフを用いて，なりすましを行う攻撃のことをウルフ攻撃と呼ぶ．通常のなりすまし攻撃が特定の攻撃対象の 1 つのデータに対して 1 つの攻撃用の入力データを用意するのと異なり，ウルフ攻撃は 1 つの入力データで複数の登録データになりすますことが可能である．

認証システムに対してウルフ攻撃が成功する確率を WAP（Wolf Attack Probability: ウルフ攻撃確率）と呼ぶ．WAP は式 (3) によって定義される．

$$\mathrm{WAP} = \frac{\text{ウルフ攻撃が成功した回数}}{\text{ウルフ攻撃の攻撃対象の数}} \times 100\ [\%] \tag{3}$$

WAP が大きいほど，対象としている認証システムがウルフ攻撃に対して弱いことがわかる．そのため，WAP は小さい値であることが望ましい．生体認証における精度の評価指標として，FRR（False Reject Rate: 本人拒否率）と FAR（False Accept Rate: 他人受入率）の 2 つが定義されており，認証システムの評価によく用いられるが，WAP は，FRR と FAR に大きく依存しないことが知られている．そのため，FRR と FAR が小さい値であるシステムにおいても WAP が高くなることが起こりうる．

4. k-means を用いたウルフ攻撃の提案手法

k-means とは，機械学習における教師なし学習の一種であり，非階層型クラスタリングを実現するためのアルゴリズムである．クラスタ数 *k* が与えられると，k-means は，クラスタリングの対象となるデータを *k* 個のクラスタに分類し，各クラスタの平均を求める．欲張り探索で局所解を求め，初期状態によって最終結果は大きく影響されるため，1 回の結果で最良のものが得られるとは限らない[5].

クラスタリング対象のデータ数を *m*，分類するクラスタ数を *k* とした時の k-means のアルゴリズムは，以下の手順 1.から手順 4.のようになる．

1. 各データ x_i（$i = 1,\ldots, m$）に対してランダムにクラスタ 1,…, *k* を割り振る．
2. 割り振ったデータをもとに各クラスタ C_j（$j = 1,\ldots, k$）の重心 V_j（$j = 1,\ldots,k$）を計算する．計算には割り当てられたデータの各要素の算術平均を使用する． x_i の *y* 番目の要素を x_{iy} とし，C_j に割り当てられたデータ数を $|C_j|$ とすると，V_j の *y* 番目の要素 V_{jy} は式 (4) で表される．

$$V_{jy} = \frac{\sum_{x_i \in C_j} x_{iy}}{|C_j|} \tag{4}$$

3. 各 x_i と各 V_j との距離を求め，x_i を最も近い重心のクラスタに割り当て直す．この距離は識別関数と同様の距離を用いる．
4. 上記の処理で全ての x_i のクラスタの割り当てが変化しなかった場合に，収束したと判断して処理を終了する．そうでない場合は新しく割り振られたクラスタから V_j を再計算して手順 1.から 3.の処理を繰り返す．

本研究ではキーストローク認証において，k-means を適用したウルフ攻撃手法を提案する．2 章で述べた本研究で取り扱うキーストローク認証は，16 次元空間におけるベクトルの距離を比較することで認証を行う．ベクトルの各要素の範囲は広く，1 章で述べた指紋認証のウルフ生成手法のように探索することでウルフを発見的に生成するのは困難である．また，同じく 1 章で述べた虹彩認証のウルフ生成手法に関しても，キーストローク認証とは，認証の方式が大きく異なるため，キーストローク認証に適応することは困難である．k-means は，*N* 次元空間におけるクラスタリングを実現するアルゴリズムであるため，登録データが密集しているクラスタを特定することができ，そのクラスタの重心ベクトルは，複数の登録者になりすますことができるウルフとなり得ることが期待される．

ウルフ攻撃を行うためには，まず攻撃側が，攻撃対象システムの登録データとは異なるキーストロークデータを独自に大量に収集する．これらを，ウルフを生成するための学習データとする．次に，k-means を用いて学習データをクラスタリングする．クラスタ数の大きいものからクラスタ 1，クラスタ 2，・・・，クラスタ *k* とする．各クラスタの重心ベクトルがウルフとなる．

許可される攻撃回数を *n* とすると，ウルフ攻撃は次のように行う．まず，攻撃対象である正規ユーザ側の登録データとクラスタ 1 のウルフとの距離を求める．2 章で述べた閾値以下であれば本人と見做され，ウルフ攻撃が成功する．ウルフ攻撃に失敗した場合，次はクラスタ 2 のウルフとの距離を求めて成功するかを調べる．以下，同様に解析を行い，クラスタ *n* までウルフ攻撃が成功するかどうかを調べる．ウルフ攻撃の評価指標としては，WAP を用いる．テストデータに対して *n* 回までの攻撃が許可されている時の WAP を WAP_n と表す．

5. 実験

5.1 実験環境

本研究では，過去の研究で開発されたキーストロークデータ収集システムによって収集された 557 人分のデータを用いる[6].

Copyright © 2019 by Information Processing Society of Japan and
The Institute of Electronics, Information and Communication Engineers
All rights reserved.

入力文字数が 500 文字であれば数分で入力でき，佐村と西村によると 1 対 N 認証で約 100 %の高い認証精度が得られている[6]．本研究では 1 対 1 認証を行い，FAR と FRR を評価した．テストデータの入力文字数が 500 文字の非定型文文書に対して 2 章のキーストローク認証手法を適用し，本人認証を 171 回，他人認証を 2028 回行った結果を表 1 に示す．他人認証と比べて本人認証の認証回数が少ないのは，本人認証では同一人物の登録データとは異なるデータのみを入力データにするためである．表 1 より，EER は *ED* を用いたとき 5.8 %程度，*MD* を用いたとき 5.3 %と，顔認証程度の高い認証精度が得られた．

以上のキーストローク認証手法を用いて，本研究ではなりすましのウルフ攻撃，さらになりすまし対策であるウルフ対策についての実証実験を行う．

表 1　識別関数ごとの FAR，FRR

識別関数	FAR [%]	FRR [%]
ED	5.72	5.85
MD	5.29	5.26

5.2 攻撃実験

機械学習における典型的な学習データとテストデータの分け方として，全被験者数 557 人のうち，30 %の 167 人分のキーストロークデータを学習データ（ウルフを生成するためのデータ）とし，残りの 70 %の 390 人分のデータをテストデータ（攻撃対象となる正規ユーザ側の登録データ）として，ウルフ攻撃実験を行う．k-means による分割を行うクラスタ数 k を変動させてウルフを生成し，それぞれの k において生成されたウルフを用いて 2 章の認証システムに対してウルフ攻撃を行う．ウルフ攻撃に用いるウルフの数 n を最大攻撃回数とする．認証システムにおいて認証拒否された際の再認証の猶予を考慮した時に 3 回程度までなら，ユーザに猶予を与えても良いだろうと判断し，$n = 3$ の時までの攻撃実験を行う．

評価指標には先述の FRR，FAR，WAP を用いる．

また，学習データから 3 つの正規ユーザのデータを抽出して，ウルフの代わりに攻撃に用いた場合の WAP_3 を $dWAP_3$ とする．ウルフ攻撃による WAP_3 との比較のために，3 つの異なる正規ユーザのデータを用いて，ウルフ攻撃と同様になりすましを試みる．この手順を 4 回繰り返し，毎回 $dWAP_3$ を算出する．毎回異なる正規ユーザのデータを 3 つずつ用いる．

5.3 攻撃実験結果

識別関数 *ED*，*MD* それぞれについて，k-means によるクラスタリングのクラスタ数 k と WAP_n の関係を攻撃実験結果として表 2，3 に示す．また，通常のデータを用いた攻撃を 4 回行った場合の $dWAP_3$ の平均値は，*ED* を識別関数に用いたとき 12.29 %，*MD* を用いたとき 12.69 %となった．

表 2，3 において，$k = 5$，160 のとき，*ED*，*MD* ともに WAP_3 は低くなった．これは，$k = 5$ の場合は，クラスタ数が少なすぎて，クラスタの領域が広くなり，1 つのクラスタに所属するデータ数が著しく多くなる．そのため，クラスタの重心が学習データの密集している場所から遠ざかりやすくなり，WAP_3 が低くなったと考えられる．また，$k = 160$ の場合は，クラスタ数がデータ数 167 個に対して多すぎるため，1 つのクラスタに所属するデータ数が著しく少なくなる．そのため，クラスタの重心が学習データの密集している場所から遠ざかりやすくなり，WAP_3 が低くなったと考えられる．

一方，$10 \leq k \leq 25$ のとき，*ED*，*MD* ともに WAP_3 が特に大きな値となった．k-means は同じクラスタ数を指定した場合でも，クラスタリングを行う度に結果が異なるため，高い WAP_3 を示している $k = 10$ や $k = 20$ がウルフ攻撃に最適であるとは本結果のみでは一概に断言できない．

ED と *MD* では，*ED* の方がより高い WAP_3 が得られた．しかし，$k \geq 25$ において，同じクラスタ数のとき，*MD* の方が WAP_3 は高くなった．EER が低くても，このように WAP_3 は高くなる可能性があるということがわかる．

本実験より，ウルフは全体の最大 31.54 %のユーザになりすますことができた．キーストローク認証においてもウルフ攻撃のリスクが大きいことが確認できる．

表 2　ウルフ攻撃結果（*ED*）

k	WAP_1 [%]	WAP_2 [%]	WAP_3 [%]
5	16.41	23.59	25.90
10	9.23	14.62	29.74
15	11.80	26.15	29.23
20	15.39	26.67	31.54
25	13.85	23.33	26.41
30	15.39	20.51	25.90
40	8.72	18.72	23.33
50	12.05	16.41	24.62
60	14.10	22.05	25.13
70	12.82	22.56	25.64
80	8.46	23.33	27.18
90	12.82	16.41	24.36
100	13.59	15.64	24.87
110	14.10	22.05	25.90
120	8.46	11.54	23.85
140	7.95	22.05	25.90
160	14.36	19.49	22.05

Copyright © 2019 by Information Processing Society of Japan and
The Institute of Electronics, Information and Communication Engineers
All rights reserved.

表 3　ウルフ攻撃結果（*MD*）

k	WAP_1 [%]	WAP_2 [%]	WAP_3 [%]
5	14.62	21.80	24.36
10	8.72	14.36	29.74
15	11.54	25.64	28.97
20	15.13	26.41	31.28
25	13.85	24.36	27.69
30	15.39	20.26	26.67
40	9.49	20.26	24.87
50	11.80	16.67	26.41
60	14.10	23.85	27.18
70	12.56	24.36	26.92
80	9.74	24.10	27.69
90	13.33	16.15	25.39
100	13.33	15.13	25.13
110	14.87	23.08	28.21
120	10.00	12.56	25.13
140	8.21	22.56	27.69
160	13.33	18.46	20.77

6. おわりに

本研究ではキーストローク認証における k-means を用いたウルフ攻撃手法の提案を行い，実証実験を行うことで危険性を確認した．k-means を用いることでキーストローク認証においてもウルフ攻撃を行うことが可能であり，通常のデータを用いて攻撃した場合の $dWAP_3$ の平均値の最大 2.6 倍程度の WAP_3 が確認され，ウルフ攻撃は通常のなりすまし攻撃より高いリスクとなり得ることを示した．また，最大の WAP_3 は *ED* がより高くなったが，$k \geq 25$ においては *MD* がより高くなり， WAP は EER に大きく依存しないことが確認できた．

今後の課題としては，k-means を用いたウルフ攻撃において，反復的にウルフを生成する場合，クラスタ数の変化がどの程度 WAP に影響を及ぼすのかを *ED*，*MD* ともに厳密に明らかにすることが考えられる．また，本研究では，k-means でクラスタリングを行ったあと，クラスタ 1 からクラスタ 3 の各代表ベクトルをウルフとして攻撃に用いたが，他のクラスタの代表ベクトルを用いた方が高い WAP を得られる可能性がある．*k* 個のクラスタの代表ベクトルの中から最適な組み合わせを選び攻撃を行ったときに，WAP がどのように変化するか調査することが考えられる．

そして，ウルフ攻撃への有用な対策手法を考え，なりすまし攻撃によるリスクを低減させることで，キーストローク認証の安全性の向上を図る必要がある．

謝辞

最後に，本研究の遂行ならびに本原稿の執筆にあたり，懇切丁寧なご指導を賜りました京都大学の吉武大地さん，明石工業高等専門学校電気情報工学科の佐村敏治教授に深く御礼申し上げます．

参考文献

[1] 大塚 玲，”バイオメトリクス認証システムのウルフ攻撃に対する安全性評価技術に関する研究”，ICT イノベーションフォーラム 2013 戦略的情報通信研究開発推進事業 (101603011) (2013).
[2] 倉橋 勇気，田中 邦明，小松 尚久，”スコア累積値を用いたキーストロークダイナミクスによる個人認証手段”，電子情報通信学会技術研究報告，CQ，2006(56)，pp. 35–40 (2006).
[3] 片岡 祥啓，宮本 貴朗，青木 茂樹，泉 正夫，福永 邦雄，”キーストロークの統計情報を利用した個人認証手法の提案”，電子情報通信学会技術研究報告，CSEC，Vol. 107，pp. 23–30 (2007).
[4] 吉武 大地，佐村 敏治，”キーストローク認証におけるウルフ攻撃”，第 8 回バイオメトリクスと認識・認証シンポジウム，S4-7 (2018).
[5] Trevor Hastie，Robert Tibshirani，Jerome Friedman，”統計的学習の基礎 —データマイニング・推論・予測—”， 共立出版 (2014).
[6] 佐村 敏治，西村 治彦，”非定型な日本語文入力におけるキーストロークダイナミクス識別”，システム制御情報学会論文誌，Vol.22，No.4， pp.145–153 (2009).

Copyright © 2019 by Information Processing Society of Japan and
The Institute of Electronics, Information and Communication Engineers
All rights reserved.