

キーストロークを用いたパスワード認証の強化について -MT システムによるパターン認証-

日大生産工 (院) ○大坂一司
日大生産工 矢野耕也

1. 概要

現在の日本において、コンピュータは情報の管理や、文章作成などの事務処理だけでなく、web サービスの利用ができるため、企業だけでなく、一般家庭にも急速に普及が進んでいる。特に、ネットショッピングや、ネットバンクなどのサービスは、その利便性から多くのユーザに利用されている。これらのサービスを利用するにあたり、本人を特定する方法として、現在は、IDとパスワードによる知識認証システムが主流であるが、パスワードは文字情報なため、外部へ漏れやすいことなどから、近年では、セキュリティ対策として、文字、数字、記号などを混ぜた長いパスワードを定期的に変更することが推奨されている。しかしながら、これらの対策はユーザに対して大きな負担を強いることから、本研究では、パスワードの入力動作に着目し、入力動作の癖や特徴から本人認証を行う方法について検討したので報告を行う。

2. キーストロークによる本人認証

キーストロークによる本人認証とは、キーストローク、すなわち、キーボードの入力操作における各人の癖や特徴から本人を特定するというものである。また、これらの特徴は、タイピングにおける各種動作時間に現れると考えられる。本研究では、 n 回目のキーが押された時の時間を Up Time (n)、離された時の時間を Down

Time (n) を測定し、これらの測定値から、 n 回目のキーが押されてから離されるまでの時間 $D_n U_n T$ (式 (1))、 n 回目のキーが離されてから次のキーが押されるまでの時間 $U_n D_{n+1} T$ (式 (2))、 k 文字の文字列を入力するためにかった時間 $D_1 U_k T$ (式 (3)) を算出し、本人を特定する特徴として用いた。

$$D_n U_n T [\text{msec}] = \text{Up Time } (n) - \text{Down Time } (n) \quad (1)$$

$$U_n D_{n+1} T [\text{msec}] = \text{Down Time } (n+1) \quad (2)$$

$$D_1 U_k T [\text{msec}] = \text{Down Time } (1) - \text{Up Time } (k) \quad (3)$$

図 1 は『AB』という文字列を入力した時の特徴量の例である。『A』を押してから離されるまでの時間が $D_1 U_1 T$ 、『B』を押してから離されるまでの時間が $D_2 U_2 T$ 、『A』が離されてから『B』が押されるまでの時間が $U_1 D_2 T$ 、『AB』と入力するのにかった時間が $D_1 U_2 T$ となる。

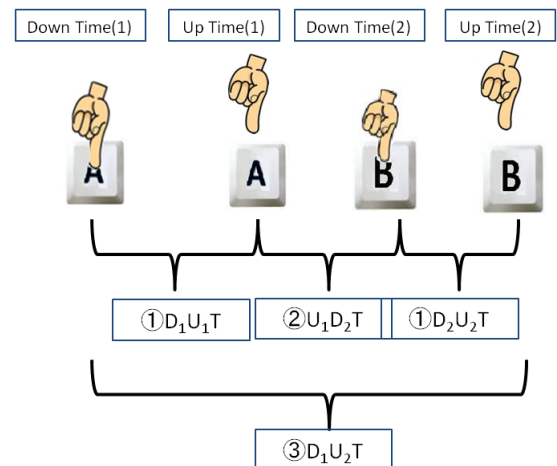


図 1 特徴量の例

About strengthening of password authentication using a keystroke

- Pattern attestation by MT system -

Hitoshi OSAKA, Koya YANO

データのサンプリングには、Windows API の Time Get Time 関数を用いて Up Time (n) 及び、Down Time (n) を測定した。Time Get Time 関数は、PC の内部時間を 1 ミリ秒単位で取得することが可能な関数で、例として『AB』と入力した際のデータを表 1、表 2 に示す。また、特徴量として D_1U_1T 、 D_2U_2T は式 (1) より

$$D_1U_1T[\text{msec}] = 78 = 2097562 - 2097484$$

$$D_2U_2T[\text{msec}] = 63 = 2097734 - 2097734$$

として算出し、 U_1D_2T は式 (2) より

$$U_1D_2T[\text{msec}] = 109 = 2097671 - 2097562$$

として算出し、また D_1U_2T は式 (3) より

$$D_1U_2T[\text{msec}] = 250 = 2097734 - 2097484$$

として算出した。

表 1 『AB』と入力した際のデータ[msec]

D_1U_1T	D_2U_2T	U_1D_2T	D_1U_2T
78	63	109	250

表 2 『AB』と入力した際の Down Time、Up Time

	Down Time[msec]	Up Time[msec]
A	2097484	2097562
B	2097671	2097734

3. 誤圧によるパターン分析

各種動作時間から本人と同定するにあたり、品質工学のパターン分析の手法の誤圧を用いた。誤圧とは、分散分析における一般平均 S_m と誤差分散 V_e を用いた合成変量で、正常状態のばらつきを基準とした単位量を単位空間として定義し、非正常状態のばらつきの程度について、定義した基準からの乖離の程度で判断するものである。識別対象データについて、望目特性の SN 比による 2 乗和の分解を行い、一般平均 S_m と誤差分散 V_e を求め、均質な集団である単位空間の分散 V_0 で除した値の和の平方根が、誤圧を利用したパターン距離となる。サンプル数に制約がなく、行列を用いていないため、計算手順が少なく、データ処理を早く行えることが特徴である。

ここでは本人のキーストローク(タイピン

グのパターン) データから単位空間を定め、誤圧の値をパターン距離とし、比較対象の人物が行った文字列の入力動作に対し、動作ばらつきの程度から本人か否かの判断を行うものである。

5. 閾値の決定方法

キーストロークによる本人認証は、動作特徴を利用したバイオメトリクスに分類されるが、動作特徴は、個人でばらつきが大きく、その特徴を的確に捉えることが困難であると言われている。そのため、図 2 のように、本人を他人と判断してしまう本人拒否や、他人を本人と認証してしまう他人受入の誤認証が発生してしまう。

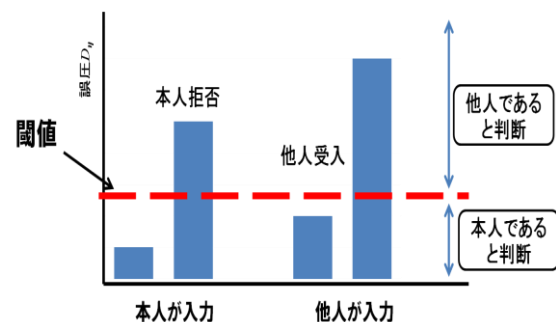


図 2 閾値と誤認証

一般的には本人拒否よりも、他人受入の方が発生した場合の損失が大きいため、閾値は他人受入のリスクを小さくする値にすることが望ましい。そこで本研究では、デジタルの標準 SN 比の解析法を応用した損益分岐点の考えを利用し、閾値を決定する。

まず本人拒否率を p 、その時の損失を C_p とする。また、他人受入率を q 、その時の損失を C_q とした時、 $p \times C_p = q \times C_q$ である場合が経済的にバランスしているため、この等式が成立する点が経済的損失の面からみて望ましいと言える。本人拒否率 p と他人受入率 q は $p = 1 - q$ で表せる確率であるが、それにより発生する損失 C_p 、 C_q はある程度現実的に発生しうる損失金額が必要なことから、ここではあるインターネットによる株取引の場面を想定した損失を求め、損益分岐点を見つ

ける方法にて閾値を決定した。

6. 実験

6.1 タイピングスキルのサンプリング

本研究では、パスワードとの併用を想定し、10 人の被験者から、『ITTEKIMASU』という文字を 100 回入力してもらい、タイピングにおける各種操作時間を測定した。また、タイピングの各種操作時間は、使用するキーボードへの慣れや、被験者のタイピングスキルが大きく関わることを考えられる。そこで、PC は同一のものを使用し、各被験者はサンプリング前に、500 文字程度のテストタイピングを行った。被験者 10 名のタイピングスキルを表 3 に示す。

表 3 被験者 10 名のタイピングスキル

	タイピングスキル	タッチタイピング	タイピングで使う指
ユーザ 1	高い	できる	全指
ユーザ 2	高い	できる	全指
ユーザ 3	普通	できない	人差し指、中指、薬指
ユーザ 4	普通	できない	人差し指、中指、薬指
ユーザ 5	普通	できない	人差し指、中指、薬指
ユーザ 6	普通	できない	人差し指、中指、薬指
ユーザ 7	普通	できない	人差し指、中指
ユーザ 8	普通	できない	人差し指、中指
ユーザ 9	低い	できない	人差し指
ユーザ 10	低い	できない	人差し指

6.2 他者へのなりすましの検証

6.1 で述べたの被験者 10 名によるサンプリングの実験では、被験者にはいつも通りの自分のペースでタイピングを行ってもらったが、キーストローク認証があるセキュリティがあると認知されていれば、不正アクセスを行う場合、他人は本人のタイピングに似せて行うことが考えられる。そこで、タイピングスキルの高いユーザ 1 がユーザ 6 に成りすますことが可能か否かを検証した。検証方法として、①～③の 3 種類の方法を行った。またサンプリング回数は各 10 回ずつ行った。

①パスワードのみを知っていた場合として、タイピングのリズムを毎回変えながら入力を行った。

②パスワードとパスワードのおおよその入力時間を知っていた場合として、指定した入力時間に合うようにタイピングを調節して行った。

③本人がパスワードを打つ様子を何回も見えて熟知していた場合として、ユーザ 6 のタイピングの様子を観察してもらい、使用する指や入力時間などを熟知した上でサンプリングを行った。

7. 結果

7.1 パターン距離による比較

ここでは、代表としてユーザ 6 を本人とした場合の各ユーザの平均パターン距離を図 3 に示した。

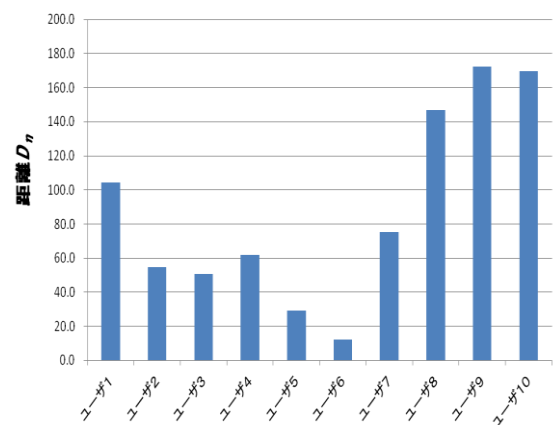


図 3 ユーザ 6 を本人とした場合の平均パターン距離

誤圧から求めたパターン距離は、タイピングスキルの違いや動作パターンの相違を示しており、数値が大きくなるほど基準とした動作から大きく異なることを示している。基準となるユーザ 6 に最も近いパターン距離を示したユーザ 5 においても、ユーザ 6 の平均パターン距離である約 15 から 2 倍以上離れているため、各ユーザの識別は可能であると判断できる。またグラフから、タイピングスキルが異なるユーザほど、パターン距離が大

きくなる傾向が見て取れた。このように、パターン距離の大きさで本人か否かを識別する可能性が示唆された。

7.2 損益分岐点による閾値の検討

閾値は 15 付近となるため、パターン距離が閾値に近いほどユーザ 6 の動作に近いと判断でき、閾値から離れるほど本人と異なるといえるが、その際、基準と同じであるか否かの識別限界である閾値が必要である。閾値の決定において必要なパラメータの決定であるが、今回はインターネットでの株取引を例にあげて行った。あるユーザが 30 万円分の株の購入するため、株取引のサイトにログインを試み、本人拒否の場合、ログインできず、株の購入ができない。この時に発生する損失 C_p は、機会損失として 6 万円発生するものとする。次に、本人が購入をする前に他人が本人になりすまし、30 万円分不正に取引を行った。この時、本人の持ち物である 30 万円と、不正発覚による捜査や、口座差止めなどの手続きなどにより 60 万円の損失 C_q が発生したとする。以上の C_p 、 C_q を仮定し、ユーザ 6 を本人として閾値であるパターン距離を 9 から 20 まで 1 ずつ変化させた場合の損益分岐点をシミュレートしたものが、表 4 に示した結果である。この結果から、パターン距離 D が 15 の時

$$p \times C_p = 0.24 \times 60,000 \text{ 円} = 14,400 \text{ 円}$$

$$q \times C_q = 0.023 \times 600,000 = 14,000 \text{ 円}$$

となる、損失がほぼバランスしている点として、この値を閾値に決定した。つまり、閾値としてパターン距離 $D=15$ を用いると、本人が拒否された場合のログインが不能による株の購入ができない場合の損失と、本人が購入する前に他人が本人になりすまして不正に取引を行った場合の損失がほぼ等しくなり、ログイン不能による期待損失と不正による損失が等しくなる。また閾値をこの値より大きくしたり小さくすると、期待損失または

不正による損失のいずれかが大きくなる。

表 4 損益分岐表

距離 D	他人受入率(q)	本人拒否率(p)	$q \times C_q$ (円)	$p \times C_p$ (円)
9	0	0.43	0	25,800
10	0.001	0.42	667	25,200
11	0.001	0.39	667	23,400
12	0.002	0.37	1,333	22,200
13	0.01	0.32	6,000	19,200
14	0.018	0.29	10,667	17,400
15	0.023	0.24	14,000	14,400
16	0.033	0.22	20,000	13,200
17	0.041	0.21	24,667	12,600
18	0.048	0.2	28,667	12,000
19	0.052	0.19	31,333	11,400
20	0.058	0.17	34,667	10,200

7.3 なりすましの検証

6.2 で行った①～③についての結果を図 4 に示す。

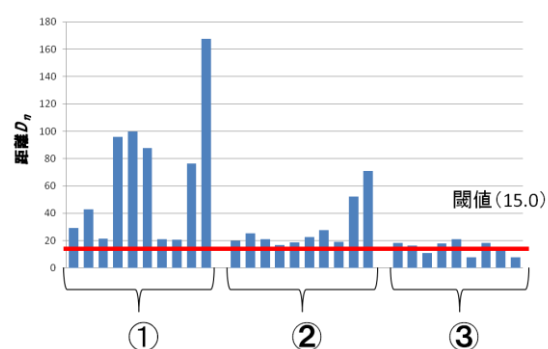


図 4 成りすましの検証結果

③のように、本人の入力時の特徴が知られた場合のパターン距離は閾値に近くなり、定型文による認証にはなりすましのリスクが観察される。

8. 結論

キーストロークのような動作の数値化で、本人認証の定量的な判断が可能であるが、なりすましなどには定型文以外の方法などが必要である。