

生体認証システムにおける新しいセキュリティ評価尺度：ウルフ攻撃確率

Wolf Attack Probability: a New Security Measure in Biometrics-Based Authentication Systems

宇根 正志 * 大塚 玲 * 今井 秀樹 *†
Masashi Une Akira Otsuka Hideki Imai

あらまし 生体認証システムにおいては、なりすましに対するセキュリティ評価尺度として、誤受入率が参照されるケースが多い。ただし、誤受入率は、複数のテンプレートに対して一致と誤判定される入力情報（ウルフと呼ばれる）を用いてなりすましを行う攻撃への耐性を評価する尺度ではない。そこで、本稿では、一致と誤判定されるテンプレート数が最大となるウルフを用いる攻撃への耐性を評価する尺度として「ウルフ攻撃確率」を提案する。

キーワード ウルフ攻撃確率、誤受入率、生体認証システム、セキュリティ評価、評価尺度

1 はじめに

近年、生体認証システムが個人を認証する技術として普及の一途を辿っており、そうした中で生体認証システムのセキュリティ評価の重要性が一層高まっている。生体認証システムに対する代表的な脅威である「第三者によるなりすまし」に関しては、その耐性の評価尺度として誤受入率（false accept rate）が用いられるケースが一般的である。誤受入率は、他人を誤って本人として受け入れてしまう確率として定義される。

しかし、複数のテンプレートと一致と誤判定される入力情報（ウルフ＜wolf＞と呼ばれる）を巧みに利用する攻撃を想定した場合、なりすましに対する耐性を誤受入率によって評価することはできない。たとえ強力なウルフが存在しても、誤受入率の測定ではサンプルの中に全体からみればわずしかウルフが存在しないため、低い誤受入率に抑えられて、ウルフの存在は一般に顕在化しない。照合アルゴリズムの性質によって強力なウルフが存在する場合には、ウルフを用いることによって誤受入率をはるかに上回る確率でなりすましに成功してしまうおそれがある。

ウルフを用いた攻撃に対する耐性を適切に評価するためには、攻撃者にとって最も有利な入力情報をアルゴリ

ズムやシステムに与えたときに得られる成功確率を求める必要がある。本稿では、こうした攻撃成功確率を「ウルフ攻撃確率（wolf attack probability）」と呼び、新しいセキュリティ評価尺度として提案する。

2 生体認証システムにおけるなりすまし

生体認証システムは、一般に、センサによる生体情報（アナログ）の取得、生体情報から特徴情報（デジタル）の抽出、予め登録されていたテンプレートとの照合（類似度の算出）、類似度に基づく判定、判定結果の出力、という一連の処理によって構成される。

こうした生体認証システムにおける代表的な脅威として、認証時における第三者によるなりすましが挙げられる。なりすましの手法は、攻撃者が人工物を用いるか否かという観点で次のように分類することができる。

・人工物を用いる攻撃

- **攻撃 A**：攻撃者は、なりすましの対象となっている個人の生体情報を取得し、それを人工物によってシステムに提示する。
- **攻撃 B**：攻撃者は、当該個人の生体情報を得ずに、適当に選択した入力情報を人工物によってシステムに提示する。

・人工物を用いない攻撃

- **攻撃 C**：攻撃者は、なりすましの対象となっている個人の生体情報を取得し、人工物以外の物体（例えば、当該個人の身体の一部）によ

* 独立行政法人産業技術総合研究所情報セキュリティ研究センター、東京都千代田区外神田 1-18-13 ダイビル 1102, Research Center for Information Security, National Institute of Advanced Industrial Science and Technology, 1102 Daibiru 1-18-13 Sotokanda Chiyoda Tokyo (masashi-une@aist.go.jp)

† 中央大学理工学部電気電子情報通信工学科、東京都文京区春日 1-13-27, Department of Electrical, Electronic, and Communication Engineering, Faculty of Science and Engineering, Chuo University, 1-13-27 Kasuga Bunkyo Tokyo

てシステムに提示する。(なりすましの対象となっている個人の生体情報を身体部分ごと分離して提示するという攻撃が含まれる.)

- **攻撃 D**：攻撃者は、当該個人の生体情報を得ずに、適当に選択した入力情報を人工物以外の物体によってシステムに提示する。(「攻撃者が自分の生体情報を提示する」といういわゆる “zero-effort attack” が含まれる.)

これらのうち、人工物を用いる攻撃は、実システムにおける現実の脅威であることが [1] をはじめとする先行研究によって示されている。このため、人工物による入力情報の提示が可能であるという前提のもとで、攻撃 A, B への耐性を評価する必要がある。攻撃 A については、攻撃者がなりすましの対象の個人の生体情報を取得できるか否かが問題となる。攻撃 B については、攻撃者がどのような入力情報を攻撃に利用するか、その結果、なりすましが成功する確率がどの程度となるかが問題となる。

攻撃 B への耐性の評価に焦点を当てると、攻撃に用いられる入力情報が、生体情報の集合から選択される場合だけでなく、生体情報としては不自然とみられる入力情報も含む、すべての入力情報の集合から選択される場合を考慮しなければならない¹。そのうえで、攻撃に最も有利な入力情報を可能な限り少ない回数提示してなりすましを行うという戦略が採用されることを前提に評価を行う必要がある。こうした攻撃として、以下のウルフ攻撃 (wolf attack) を定義する。

【定義】ウルフ攻撃：攻撃者が、すべての入力情報の集合の中から、一致と誤判定されるテンプレート数が最大となる入力情報 (ウルフ) を探索し、そのウルフを提示することによってなりすましを試みる。

3 既存の評価尺度：誤受入率

誤受入率は次のように定義される。

【定義】誤受入率 (FAR)：システムへの入力情報を s 、生体情報の集合を S_h 、(s に対応しない) テンプレートを t 、同システムに登録されている (人間の生体情報から生成された) テンプレートの集合を T_h 、 s と t を与えたときにシステムが出力する判定結果を $match(s, t) \in \{accept, reject\}$ とすると、

$$FAR \triangleq \text{Ave}_{t \in T_h} \text{Ave}_{s \in S_h} Pr[match(s, t) = accept]. \quad (1)$$

“ $Pr[match(s, t) = accept]$ ” は s と t が一致と誤判定される確率であり、“ $\text{Ave}_{t \in T_h} X$ ” は t を変化させたときの X の平均値を表している。

¹ 「生体情報してありえない人工的な入力情報を受け入れてしまう」という脆弱性は “synthesised biometric samples” と呼ばれる脆弱性として知られている。

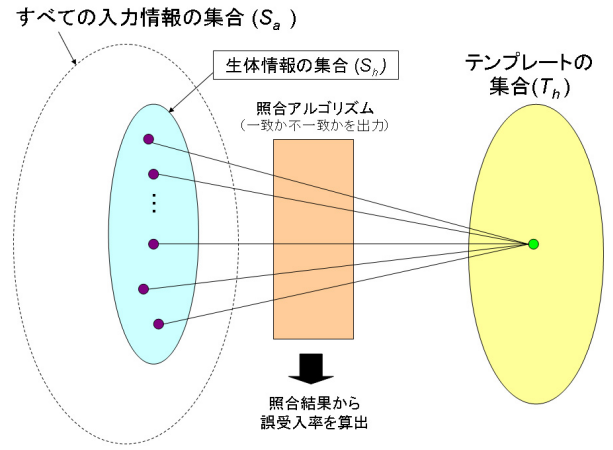


図 1: 誤受入率の測定

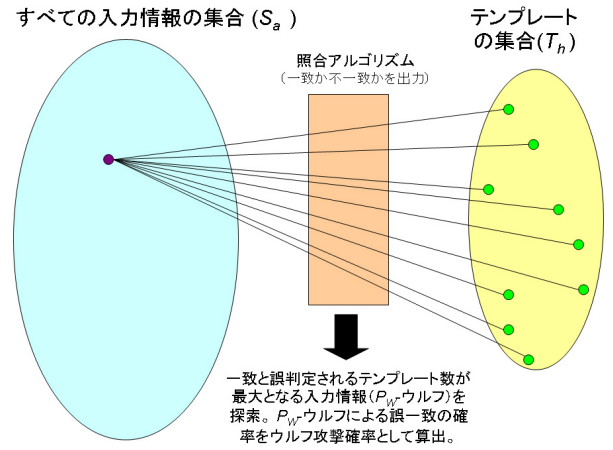


図 2: ウルフ攻撃確率の測定

このように、誤受入率の測定では、テンプレートを固定したうえで各生体情報に対して一致と誤判定される確率の平均を計算し、次にテンプレートを変えて同様の確率を計算するという形となっている。このため、たとえ強力なウルフが入力情報の集合に存在したとしても、集合全体からみればわずかな数となり、低い誤受入率に抑えられて、ウルフの存在は一般に顕在化しなくなってしまう。このように、誤受入率はウルフ攻撃への耐性を表す尺度ではない。

4 新しい評価尺度：ウルフ攻撃確率

ウルフ攻撃への耐性を評価する尺度として、以下のウルフ攻撃確率を提案する。

【定義】ウルフ攻撃確率 (P_w)：

$$P_w \triangleq \max_{s \in S_a} \text{Ave}_{t \in T_h} Pr[match(s, t) = accept]. \quad (2)$$

ウルフ攻撃確率 P_w を生じさせる入力情報を「 P_w -ウルフ」と呼ぶ。 P_w -ウルフは、一致と誤判定されるテンプレートの数が最大となる入力情報であり、 S_a の中から探索される (図 2 参照)。ウルフ攻撃では、攻撃者は、 P_w -ウルフを入力情報として $1/P_w$ 回提示することによって、任意のテンプレートに対して一致と誤判定させることが可能となる。

5 ウルフ攻撃のアイデアに基づく評価例

ウルフ攻撃のアイデアに基づくセキュリティ評価研究の事例として、渡邉らの研究 [2] と河上らの研究 [3] を紹介する。これらの研究は、ウルフ攻撃確率が誤受入率よりも有意に大きな値となる可能性を示している。

5.1 指静脈パターン照合アルゴリズムへの適用例

渡邉らは、三浦ら [4] の指静脈パターン照合アルゴリズムにおいて 1-ウルフ (ユニバーサル・ウルフと呼ばれている) が存在することを示し、ユニバーサル・ウルフを 1 回提示することによって任意のテンプレートに対して一致と誤判定させることが理論的に可能であるとしている。一方、三浦らの論文 [4] には、678 人の被験者から採取された指静脈パターンによる誤受入率の計測結果が示されており、誤受入率と誤拒否率の差が最小となる場合の誤受入率として 0.145 % が得られたとしている。

5.2 マニューシャ照合アルゴリズムへの適用例

河上らは、指紋認証で広く利用されている特徴点 (マニューシャと呼ばれる) を用いた照合アルゴリズムを対象に、ランダムに選択した入力情報とテンプレートにおいて一定数以上の特徴点が一致する確率 (マニューシャ衝突率) を求め、マニューシャ衝突率を最大化する入力情報の探索を行っている。こうした河上らの検討は Ratha らの検討結果 [5] をベースとしている。Ratha らは、入力情報とテンプレートのサイズが同一のケースのみを対象にマニューシャ衝突率を導出している。一方、河上らは、すべての入力情報の集合から最も攻撃に有利な入力情報を選択する場合のマニューシャ衝突率を導出している。こうした攻撃はウルフ攻撃に対応する。

Ratha らは、入力情報とテンプレートの特徴点数を 40、入力データに含まれる最大の特徴点数を 400、特徴点の方向情報のバリエーションを 4 に設定し、10 以上の特徴点が一致する場合のマニューシャ衝突率を約 2^{-26} と見積もっている。一方、河上らは、特徴点数が最大値 400 となる入力情報のときにマニューシャ衝突率が最大になることを示し、10 以上の特徴点が一致する場合のマニューシャ衝突率を約 0.6 と見積もっている。

6 おわりに

本稿では、まず、なりすましに対する耐性を評価する尺度として誤受入率が適切とはいえないこと、および、一致と誤判定されるテンプレート数が最大になる入力情報を用いる攻撃 (ウルフ攻撃) を想定して評価する必要があることを説明した。そのうえで、ウルフ攻撃に対する耐性を評価する尺度としてウルフ攻撃確率を提案し、本尺度の基づく既存の評価研究結果を説明した。今後、生体認証システムにおけるなりすましへの耐性を評価する際には、ウルフ攻撃確率を算出することが必要である。

参考文献

- [1] T. Matsumoto, H. Matsumoto, K. Yamada and S. Hoshino, "Impact of Artificial "Gummy" Fingers on Fingerprint Systems," Optical Security and Counterfeit Deterrence Techniques IV, Proceedings of SPIE, Vol. 4677, pp.275-289, 2002.
- [2] 渡邉直彦・繁富利恵・宇根正志・大塚玲・今井秀樹, "指静脈パターン照合アルゴリズムにおけるユニバーサル・ウルフ," コンピュータセキュリティシンポジウム 2006 予稿集, 情報処理学会, pp.621-626, 2006 年
- [3] 河上梨恵・繁富利恵・美添一樹・宇根正志・大塚玲・今井秀樹, "マニューシャ・マッチングのウルフに関する理論的考察," 2007 年暗号と情報セキュリティ・シンポジウム予稿集, 電子情報通信学会, 2007 年
- [4] 三浦直人・長坂晃朗・宮武孝文, "線追跡の反復試行に基づく指静脈パターンの抽出と個人認証への応用," 信学論 D-II, Vol. J86-D-II, No. 5, pp.678-687, 2003 年
- [5] N. K. Ratha, J. H. Connell and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," IBM System Journal, Vol.40, No.3, pp.614-634, 2001.
- [6] 渡邉直彦・繁富利恵・美添一樹・宇根正志・大塚玲・今井秀樹, "指静脈パターン照合アルゴリズムにおけるユニバーサル・ウルフ —特徴抽出過程を含めた考察—," 2007 年暗号と情報セキュリティ・シンポジウム予稿集, 電子情報通信学会, 2007 年