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（101603011）

Theoretical Security Evaluation of Biometric Authentication Systems against Wolf Attacks

研究代表者
大塚　玲　産業技術総合研究所
Akira Otsuka　AIST

研究分担者
井沼 学[†]　今井秀樹[††]　甲藤二郎[†††]　大木哲史[†††]
Manabu Inuma[†]　Hideki Imai[††]　Jiro Katto[†††]　Tetsushi Ohki[†††]
[†]産業技術総合研究所　[††]中央大学 理工学部　[†††]早稲田大学 理工学術院
[†]AIST　[††]Chuo University　[†††]Waseda University

研究期間　平成 22 年度～平成 24 年度

概要

バイオメトリクス認証に対する意図的ななりすまし攻撃であるウルフ攻撃とその安全性評価尺度であるウルフ攻撃確率の理論を、生体検知システムやマルチモーダル認証システムなどのより高度で実用的なシステムにも応用可能な理論に発展させ、種々のバイオメトリクス認証システムへの意図的な成りすまし攻撃に対する安全性評価理論を構築した。また、脆弱性の調査や実験的な評価手法の研究を行い、評価技術の基盤となるデータや知見を蓄積した。さらに、ウルフ攻撃に対して理論的に安全な照合アルゴリズムを開発した。

１．まえがき

バイオメトリクスは、高い利便性を達成できる重要な本人確認の手段として、空港での出入国管理、銀行 ATM、入退管理など、幅広く利用されているが、安全性評価技術や評価基準が確立していないため、十分に安心して利用できるだけの認証基盤とはなり得ていない。国民が安心して利用できる安全なバイオメトリクス認証システムの普及のためには、安全性評価理論・評価技術の研究開発が重要かつ急務な課題である。

本研究開発では、認証システムのセキュリティにおいて最も重要な成りすまし攻撃に対するセキュリティ評価技術の研究開発を行った。従来のバイオメトリクス認証技術は、成りすまし攻撃に対するセキュリティ評価尺度として認証精度の評価尺度である他人受入率（FAR）を採用するのが一般的であった。しかし、FAR は利用者が通常の環境で使用する、あるいは、攻撃者が自身の生体サンプルを提示して成りすまそうとする場合の誤一致率であり、攻撃者が誤一致を起こしやすいサンプルを提示するなどして意図的に成りすましを行うときのセキュリティ評価尺度として十分ではない。そこで、本研究では、図 1 に示すような特殊な（人工的な）サンプルの提示も含めた自由度の高い攻撃を想定し、数多くのユーザのテンプレート(登録生体情報)と高い類似度を示す人工的なサンプルを提示する、より強力なりすまし攻撃（ウルフ攻撃）について研究を行った。研究代表者らは、ウルフ攻撃に対する安全性評価尺度として、あらゆるウルフ攻撃の成功確率の最大値であるウルフ攻撃確率（WAP）を提案し、指紋や静脈などの主要なモダリティ（生体情報の種類）の典型的な照合アルゴリズムに対して WAP（の下界）を示し、これまでの FAR では捉えられなかった脆弱性を発見することに成功した。さらに、理論的にウルフ攻撃に対して安全でかつ高性能な照合アルゴリズムを開発した。

２．研究開発内容及び成果

① ウルフ攻撃安全性評価理論の体系化

指紋、虹彩、静脈、話者認証など実用化されている各種認証アルゴリズムに対して、読取センサを含むサンプル取得・特徴抽出アルゴリズムに対する攻撃モデルの構築、照合アルゴリズムに対する攻撃モデル、セキュリティ評価手法を検討し、ウルフ攻撃セキュリティ評価理論の体系化を行った。

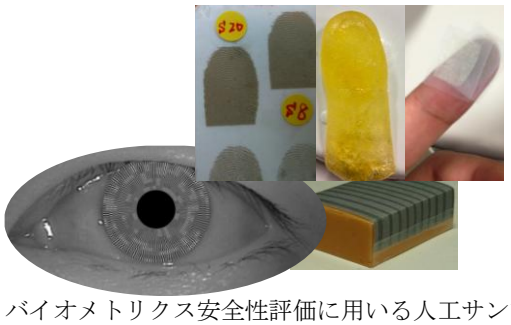

図 1 バイオメトリクス安全性評価に用いる人工サンプル

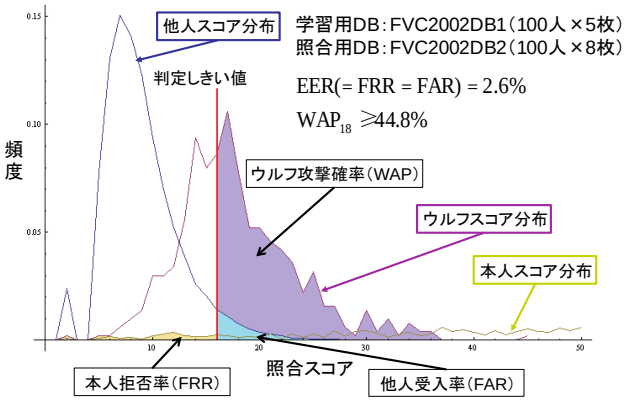

図 2 マニューシャリレーション方式の指紋照合アルゴリズムに対するウルフ攻撃実験の照合スコア分布

特に、指紋認証で一般的なマニューシャリレーション方式の指紋照合アルゴリズムに対して、米国の National Institute of Standards and Technology（NIST）が公開する照合アルゴリズム Bozorth3 を対象としてウルフ攻撃モデルを提案し、図 2 に示すような高い誤一致確率を有するウルフ指紋パターンの発見に至った。実験により EER（他人受入率 FAR と本人拒否率 FRR が等しくなるときの値）2.6%を示す閾値を設定した際、ウルフ攻撃確率（WAP）（の下界）が 44.8%と大変大きな値となることを示した。

また、話者認識アルゴリズムの安全性評価として、偽声対策を施した話者照合手法として知られている Reynolds らの手法に対して解析を行った。この手法では、予め多人数の音声データから詐称者モデル（Universal Background Model; UBM）と呼ばれるテンプレート情報を作成し、照合の際は個人テンプレートだけではなくこの UBM とも比較することで照合を行っている。図 3 に示すように、論文[2]に示したウルフ攻撃を行うことで、よく知られている LPC を用いた話者認証システムであれば、UBM を用いても 50%以上の確率でウルフ攻撃が可能となることを示した。

□② 安全な照合アルゴリズムの開発

安全な照合アルゴリズムの構成方法のフレームワークの提案と、各種モダリティに対してそれらを適用した効率のよいアルゴリズムの提案を行った。

論文[1]で提案しているアルゴリズムは、入力特徴データと人間の特徴データの照合スコア分布を推定して判定しきい値を決めることで、ウルフ攻撃による誤一致確率（なりすまし成功確率）を理論的に一定値以下に抑えることを可能にしている。しかし、安全性を完全に達成するには、入力特徴データに応じて照合スコア分布を厳密に求める必要があるため、非常に大きな計算量を要する。そこで、ある程度の安全性を保ちつつ実用に耐える効率的な認証アルゴリズムの実現には、高精度かつ高速な照合スコア分布の推定法が必要である。提示されたサンプルから算出されるスコア分布推定ための有益な指標を本研究ではクオリティと呼ぶ。例えば、J.Daugman が提案する照合アルゴリズムにおいて、瞼やまつげの影響を受けない有効ビットの長さに応じて照合スコアを正規化する手法を提案しているが、このとき有効ビット長は特徴データから算出されるクオリティとみなすことができる。今年度は、指紋認証におけるクオリティとして NIST が提案する NIST Fingerprint Image Quality（NFIQ）に注目し、これを用いた効率の良い照合アルゴリズムを提案した。

３．今後の研究開発成果の展開及び波及効果創出への取り組み

本研究成果をもとに、より本格的な安全性評価技術の研究開発を進め、国際標準および認証制度等の開発に貢献する予定である。特に、現在、ISO/IEC JTC1 SC37 において、成りすまし行為を検出する機能（bPAD）の安全性評価に関する規格（ISO/IEC 30107: bPAD-Biometric Presentation Attack Detection）が進行中であり、これを含め、海外機関と連携しつつ国際標準規格の開発に貢献を行う予定である。

４．むすび

本研究を実施したことにより、バイオメトリクス認証システムのなりすまし攻撃の大きな脅威であるウルフ攻撃に対する安全性評価の理論および実際的な評価方法の体系化を大きく進めることができた。今後、本研究の成果を国際標準の開発や認証制度の整備に活かし、安心して利用できるバイオメトリクス認証システムの普及に貢献したいと考えている。

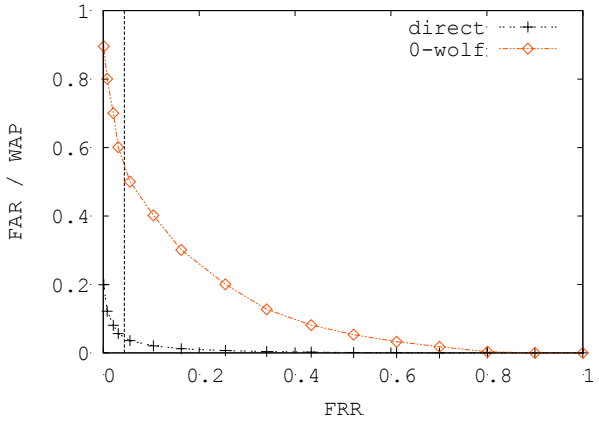

図 3 LPC 話者認証アルゴリズムに対するウルフ攻撃

【誌上発表リスト】

[1] M.Inuma, A. Otsuka and H. Imai、“A theoretical framework for constructing matching algorithms secure against wolf attack”、IEICE Transactions on Information and Systems Vol. E96-D, No.2, pp.357-364（2013 年 2 月 1 日）

[2] Tetsushi Ohki, Seira Hidano and Tatsuya Takehisa、“Evaluation of Wolf Attack for Classified Target on Speaker Verification Systems”、International Conference on Control, Automation, Robotics and Vision(ICARCV2012),（2012 年 12 月 5 日）

[3]竹久 達也、大木 哲史、井沼 学、“指紋認証システムに対する目標分類型ウルフ攻撃手法とその評価”、第 4 回 21 世紀科学と人間シンポジウム論文誌（第 4 巻）pp.20-25（2011 年 3 月 30 日）

【受賞リスト】

[1] 篠永 崇史、2010 年度「研究と実務融合による高度情報セキュリティ人材育成プログラム」(ISS スクエア) シンポジウム研究奨励賞、“Study on Vulnerability of Quality-Based Fusion Schemes（クォリティ値を考慮に入れたフュージョンスキームの脆弱性に対する考察）”、2011 年 3 月 1 日

[2] 中村 拓也、2011 年度「研究と実務融合による高度情報セキュリティ人材育成プログラム」(ISS スクエア) シンポジウム研究奨励賞、“Wolf Attacks on Multibiometrics Authentication systems（マルチバイオメトリクス認証に対するウルフ攻撃）”、 2012 年 3 月 2 日

[3] 中村拓也、2012 年度「研究と実務融合による高度情報セキュリティ人材育成プログラム」(ISS スクエア) シンポジウム準 ISS スクエア賞、“高度な指紋認証に対するウルフ攻撃（Wolf Attack against Advanced Fingerprint Recognition）”、 2013 年 3 月 1 日