P-37

情報通信技術

ディープラーニングによる セキュリティ回路 PUF への攻撃とその対策

TECHNOLOGY SHOWCASE

SATテクノロジー・ショーケース2020

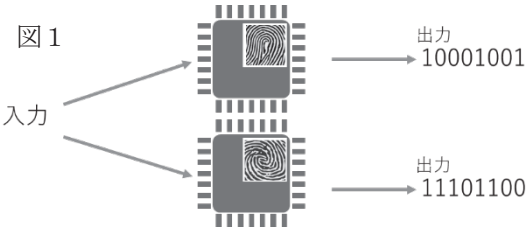
■ 背景：IoTの普及とセキュリティの課題

IoT技術の普及により，様々なデバイスがインターネットに接続され通信を行なっている．これにより，日常生活の質が向上する一方で，IoTデバイスのセキュリティリスクが増大している．特に近年，ICチップの偽造品が流通しており問題となっている．偽造品のICチップが流通することで，意図しない機密情報の流出や品質低下，不良品の流通によるブランドイメージの失墜等が起きる．

■ Physically Unclonable Function (PUF)

ICチップの真正性を確認する技術としてPhysical Unclonable Function (PUF) が注目されている (図1)．PUFは半導体デバイスのばらつきを利用してチップに固有の値を生成する回路である．同じ設計の回路に同じ入力(＝チャレンジ)を与えても，ばらつきの影響で異なる出力(＝レスポンス)となる．ゆえに，PUFを物理的に複製することは極めて困難であり，この特長を生かしてチップ固有の鍵生成やデバイス認証に用いることができる．

図1



■ 新たな課題：PUFに対する機械学習攻撃

典型的なPUFの1つであるArbiter PUFでは、攻撃者がチャレンジとレスポンスのペアを複数入手しこれを機械学習することで，正規のPUFと似た挙動を示すクローンを作製できることが報告されている。このクローンは正規PUFになりすますことができるため，PUFの実用上の脅威となり得る。機械学習攻撃を困難にするために，n個のArbiter PUFのレスポンスをXORする*n*-XOR PUFも提案されているが，機械学習技術の発展やレスポンス取得方法の工夫により，攻撃可能であることが示されている．

本研究では特に強力な機械学習であるDeep Learningを使用し，*n*-XOR PUF (n=2～5) の脆弱性評価を行った．また，対策として，レスポンスにノイズを混入させることで攻撃を困難にする手法について検討を行った．

■ 研究内容

1. Deep Learningを用いた*n*-XOR PUFの機械学習攻撃

n=2～5の*n*-XOR PUFをそれぞれ5個ずつ作成し，Deep Learning攻撃への耐性を評価した．まず，先行研究で報告された非常に小規模なDeep Learningモデルを用いて*n*-XOR PUFのクローンを試みた．その結果，n=2～4の*n*-XOR PUFのレスポンスの95%以上の予測に成功した．一方，5-XOR PUFのうち2個はレスポンスを予測できなかった．この結果，*n*-XOR PUFは*n*の増加に伴い攻撃耐性が向上するものの， Deep Learningの利用によって攻撃が可能であることがわかった

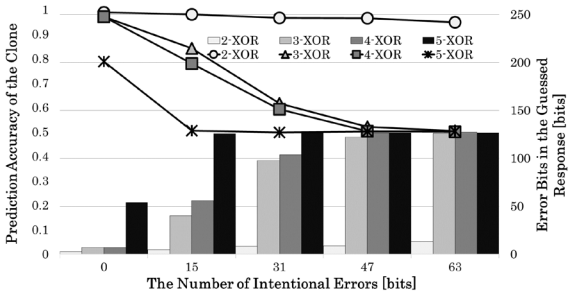
2. 意図的なノイズ混入による機械学習攻撃への対策

レスポンスに意図的にノイズを混入することによる機械学習攻撃の予測精度の低減効果を評価した．

先行研究の小規模なDeep Learningモデルでは，ノイズを入れることでレスポンスの予測精度が大きく下がり，提案手法の有効性を確認することができた (図2)．

一方，本研究で大規模なDeep Learningモデルを用いたところ，n=2～4の*n*-XOR PUFではノイズを25%混入しても予測精度は低減できなかった．しかし，5-XOR PUFではノイズを20%ほど混入すれば予測精度が下がり，攻撃者によるクローンの作製を防ぐことができることが分かった．

図2



■ まとめ

n-XOR PUFに対してDeep Learningを用いた機械学習攻撃を行い，耐性を評価した．また，レスポンスにノイズを混入し，クローンの予測精度の低減効果を評価した．小規模モデルではノイズ混入による予測精度の低減効果が非常に大きかった。大規模モデルの下では，*n*が小さい場合，ノイズを25%混入させても攻撃を防ぐことはできず，Deep Learning攻撃の強力さが確認された．

この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）の委託事業の結果得られたものです。

代表発表者 **八代 理紗（やしろ りさ）**
所　　属 **電気通信大学大学院 情報理工学研究科**
情報学専攻 崎山研究室
問合せ先 **〒182-8585 東京都調布市調布ヶ丘 1-5-1 E3-720**
TEL : 042-443-5762 FAX : 042-443-5257
yashiro@uec.ac.jp

■キーワード: (1)ハードウェアセキュリティ
(2)PUF
(3)機械学習攻撃

■共同研究者: 堀 洋平
所属 国立研究開発法人 産業技術総合研究所
ナノエレクトロニクス研究部門
エレクトロインフォマティクスグループ