SCIS 2022 2022 Symposium on
Cryptography and Information Security
Osaka, Japan & Online, Jan. 18 – 21, 2022
The Institute of Electronics,
Information and Communication Engineers

敵対的サンプル攻撃を適用した回路設計情報における ニューラルネットワークを用いたハードウェアトロイ識別に関する特徴量の検討

Effective Feature Extraction Against Adversarial Example Attacks in Hardware-Trojan Detection at Gate-Level Netlists

加藤 友浩 * Tomohiro Kato
山下 一樹 † Kazuki Yamashita
長谷川 健人 ‡ Kento Hasegawa
披田野 清良 ‡ Seira Hidano
清本 晋作 ‡ Shinsaku kiyomoto
戸川 望 † Nozomu Togawa

キーワード ハードウェアトロイ，ゲートレベルネットリスト，機械学習，ニューラルネットワーク，敵対的サンプル

あらまし

近年，IC の設計・製造の外部への委託の増加に伴い，悪意のある機能を持った回路であるハードウェアトロイが挿入される可能性が指摘されている．この脅威への対策手法として，回路設計情報から特徴量を抽出し，ニューラルネットワークを用いてハードウェアトロイを識別する手法が提案されている．この手法では，ハードウェアトロイの特徴となりうる 51 個の特徴量のうち，特に有効な 11 個の特徴量を用いて識別をしている．一方，回路設計情報に論理的に等価な改変を加え，ハードウェアトロイ識別の精度低下を促す敵対的サンプル攻撃が登場している．本稿では，上記 51 個の特徴量に加え，ならびにハードウェアトロイのトリガ回路の特徴となる 25 個の特徴量を合わせた 76 個の特徴量の内，回路設計情報への敵対的サンプル攻撃に対して堅牢性が高い特徴量を検討する．検討した特徴量を用いた識別器を，従来の 51 個，11 個，36 個（上記 11 個と 25 個の特徴量を合わせた特徴量），76 個の特徴量を用いた識別器それぞれと比較した結果，敵対的サンプル攻撃を適用した回路における識別精度に対して，堅牢性の高いことを確認した．

* 早稲田大学基幹理工学部情報通信学科, 〒 169-8555 東京都新宿区大久保 3-4-1. Dept. Communications and Computer Engineering, Waseda University, 3-4-1, Ookubo, Shinjuku-ku, Tokyo, 169-8555, JAPAN. tomohiro.kato@togawa.cs.waseda.ac.jp
† 早稲田大学大学院基幹理工学研究科情報理工・情報通信専攻, 〒 169-8555 東京都新宿区大久保 3-4-1. Dept. Computer Science and Communications Engineering, Waseda University, 3-4-1, Ookubo, Shinjuku-ku, Tokyo, 169-8555, JAPAN.
‡ 株式会社 KDDI 総合研究所, 〒 356-8502 埼玉県ふじみ野市大原 2-1-15. KDDI Research, Inc., 2-1-15, Ohara, Fujimino-shi, Saitama, 356-8502, JAPAN.