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

# グラフ学習にもとづく不正回路検知に対する強化学習を用いた自律的な脆弱性検査の提案

長谷川 健人 * 披田野 清良 * 福島 和英 *
Kento Hasegawa Seira Hidano Kazuhide Fukushima

**キーワード** AI セキュリティ, 機械学習, 強化学習, グラフ学習, ハードウェアセキュリティ

## あらまし

Society 5.0 では多種多様なソフトウェアや端末がネットワークを介して相互に接続されるため，安心・安全なデータ処理・通信には効率的なセキュリティ技術が必要となる．強化学習を用いた自律的なセキュリティ技術 [1] の活用が期待されるが，特定のアプリケーションにどのように適用するか検討する必要がある．そこで本稿では，強化学習を用いた自律的なセキュリティ技術に着目し，それを具体的なアプリケーションへ実際に適用する．

近年のセキュリティ脅威の一つに，サプライチェーン複雑化に伴う IC 設計情報への不正回路挿入が指摘されている．機械学習にもとづく不正回路検知が研究されており，直近ではグラフ学習を用いた手法 [2] も提案されている．しかし，検知が難しい不正回路の検討が不十分である．そのため，本稿で具体的なアプリケーションとして不正回路検知を扱い，検知が難しい不正回路を見つけることで，検知手法の脆弱な部分を明らかにする．

本稿では，強化学習を用いた自律的な試行により，検知システムの脆弱性を効率的に検査する手法を提案する．具体的には，グラフ学習を用いた不正回路検知手法に対して強化学習を用いた自律的な試行を適用することで，その検知を回避する不正回路サンプルを効率的に生成する手法を構築する．図 1 に提案手法の概要を示す．提案手法の利用者は，不正回路検知手法の開発者またはその利用者であり，不正回路検知の内部情報にアクセス可能である．強化学習アルゴリズムは，グラフ学習を用いた不正回路検知における判定結果と，グラフ学習で抽出された特徴ベクトルを観測する．観測された特徴ベクトルをもとに，不正回路を構成するためのパラメータを変更し，新たな不正回路サンプルを生成する．一連の操作を

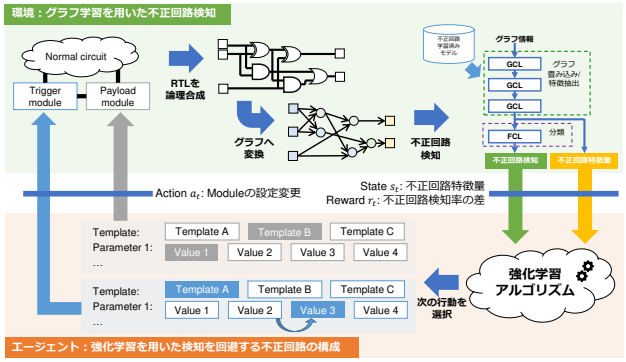


図 1: 提案手法の概要．

繰り返しながら，グラフ学習にもとづく不正回路検知において検知精度を最小化するよう，不正回路を構成するためのパラメータ変更の方策を学習する．評価実験を通じ，提案手法はランダムにサンプルを生成する方法と比較して，検知精度を低下させる不正回路のサンプルをより効率的に生成することを確認した．また，一連の実験を通じ，強化学習を用いた自律的なセキュリティ技術をグラフ学習にもとづく不正回路検知に適用する方法や課題，および，提案手法で生成された不正回路のサンプルから対象とした検知手法で検知が困難な不正回路の構成を考察した．

## 参考文献

[1] T. T. Nguyen and V. J. Reddi, “Deep reinforcement learning for cyber security,” 2019. [Online]. Available: http://arxiv.org/abs/1906.05799v3

[2] R. Yasaei, S.-Y. Yu, and M. A. A. Faruque, “Gnn4tj: Graph neural networks for hardware trojan detection at register transfer level,” in *2021 Design, Automation & Test in Europe Conference Exhibition (DATE)*, 2021, pp. 1504–1509.

* 株式会社 KDDI 総合研究所, 埼玉県ふじみ野市大原 2-1-15, KDDI Research, Inc., 2-1-15, Ohara, Fujimino-city, Saitama