早稲田大学大学院 基幹理工学研究科

博士論文審査報告書

論　文　題　目

Hardware-Trojan Detection Methods Utilizing Machine Learning Based on Hardware-Specific Features

ハードウェア固有の特徴にもとづく機械学習を利用したハードウェアトロイ検出

申　請　者

Kento	HASEGAWA
長谷川	健人

情報理工・情報通信専攻 情報システム設計研究

2020 年　2 月

近年，IoT（Internet-of-Things）デバイスが急速に普及し，近い将来，全世界で数百億個の IoT デバイスがネットワークに接続されると言われる．このような IoT デバイス，特に，IoT デバイスの基幹部品たる半導体集積回路は，主に設計工程と製造工程を経て市場に供給されるが，設計製造の効率化・安定供給などを目的に，これらのうち多くの工程が，海外の設計製造メーカをはじめ，自社以外の第三者に委託される場合が多い．さらに，自社で半導体集積回路が設計製造される場合においても，第三者によって設計製造された既設計部品（Intellectual Property; IP と呼ばれる）や設計ライブラリがしばしば使用される．その一方，半導体集積回路のサプライチェーンはさまざまなセキュリティ上の危険にさらされていると言われ，第三者の設計製造メーカにより悪意のある設計製造情報の改変が行われる可能性も指摘されている．

ハードウェアトロイとは，第三者によって，悪意を持って集積回路に埋め込まれた不正回路や不正部品の総称であり，集積回路の設計工程あるいは製造工程，いずれにおいても，ハードウェアトロイ挿入の危険性が指摘されている．**集積回路に埋め込まれたハードウェアトロイの検出は，これまでテスト設計を応用する技術，ハードウェアトロイが含まれていないゴールデンデータやゴールデンチップと設計データや製造チップとを比較することでハードウェアトロイを検出する技術，ハードウェアトロイの特徴とパタンマッチングすることで集積回路設計データ中のハードウェアトロイを検出する技術等が提案されてきた．**いずれの技術も，設計製造された集積回路から既知のハードウェアトロイを検出することに注目したものである．一方，ハードウェアトロイそのものとハードウェアトロイ検出はともに進化を続け，未知なるハードウェアトロイ検出は，これまで最大の課題となっていた．

以上のような背景ならびに議論のもと，本論文は，集積回路のハードウェアトロイ検出に対し，機械学習を応用するブレークスルーを目指し，次の 2 点を提案している：第一に集積回路の設計工程に焦点を当て，未知のハードウェアトロイを含む集積回路設計データを対象に，集積回路設計データ固有の特徴を学習し，集積回路設計データを構成する信号線を，ハードウェアトロイを構成する信号線（ハードウェアトロイ信号線）と，通常回路を構成する信号線（通常信号線）とに分類することを可能としたハードウェアトロイ検出手法を提案している．第二に集積回路の製造工程に焦点を当て，消費電力の観点で製造後の集積回路の動作を観測し，正常動作と異常動作とを分類することを可能とした異常動作検出手法を提案している．そして，それぞれの手法についてベンチマーク回路やプロトタイプ回路設計を通して，手法の有効性を評価している．

本論文は 5 章から構成される．以下では，各章の概要を述べ，評価を加える．

第 1 章「Introduction」では，本論文の背景と目的および概要をまとめ，著者の研究の位置付けを明らかにしている．

第 2 章「Hardware Trojan Classification Utilizing Machine Learning」では，集積回路の設計工程に焦点を当て，まず，機械学習によりハードウェアトロイ信号線と通常信号線とが識別できることを示すため，ハードウェアトロイを構成する回路全体の入力数が，通常回路と比較して，過度に大きくなっていることや，プライマリ入力・プライマリ出力に近いことなどを利用して，ハードウェアトロイ信号線と通常信号線とを有意に識別する 5 つのハードウェアトロイ特徴量を提案している．実際に，識別器にサポートベクタマシンを用いた実験を通して，提案した 5 個の特徴量によりハードウェアトロイ信号線と通常信号線との識別が可能であることを示している．続いて，さらに正確なハードウェアトロイ信号線の識別のため広範囲の特徴量の最適設計を行い，特徴量の寄与度を定量的に評価することで 11 個のハードウェアトロイ特徴量を提案している．実際に，識別器にランダムフォレストを用いることで，提案した 11 個の特徴量により，高い精度でハードウェアトロイ信号線と通常信号線との識別が可能であることを示している．

第 3 章「Application of the Hardware-Trojan Detection Utilizing Machine Learning」では，第 2 章で提案した 11 個のハードウェアトロイ特徴量を用いて，多段ニューラルネットワークの最適化を行い，ハードウェアトロイ信号線と通常信号線とを識別している．交差検証の結果，TPR（True Positive Rate; 全ハードウェアトロイ信号線のうち，ハードウェアトロイ信号線として検出されたものの割合）は 84.8%に達し，実利用を見据えて，十分な識別結果を得るに至っている．さらに，ハードウェアトロイの局所性を利用して，ハードウェアトロイ信号線と識別された信号線の周辺にある信号線の状態を利用することで，ハードウェアトロイ検出の改良も行っている．

第 4 章「Malicious Behavior Detection Based on Power Analysis」では，集積回路の製造工程に焦点を当て，製造後の集積回路の電力を観測することで，回路の正常動作とハードウェアトロイ等による異常動作とを識別する手法を提案している．まず回路動作をモデル化し，モデル上で電力を観測する．観測した電力情報から，正常動作と異常動作とを識別するものである．実際に，IoT デバイスとして，組込みマイクロコントローラを取り上げ，その動作をモデル化し，電力を測定した結果，正常動作と異常動作との識別に成功している．

第 5 章「Conclusion」では，本論文を総括している．

以上が本論文の概要であるが，本論文は，世界に先駆けて，集積回路に埋め込まれたハードウェアトロイ検出に機械学習を応用する手法を提案したもので，学術的に新たな領域を開拓したと同時に，ハードウェアトロイ検出の実用化に向けた扉を開いたものと位置づけられる．これらの成果は，高度情報通信社会を支える重要な基盤情報技術たるハードウェアセキュリティ技術の発展に寄与するところが大である．よって本論文は博士（工学）早稲田大学の学位論文として価値あるものと認める．

2020 年 2 月

審査員　主査	早稲田大学教授	博士(工学)早稲田大学	戸川　望
	早稲田大学教授	工学博士(早稲田大学)	柳澤政生
	早稲田大学教授	博士(情報科学)早稲田大学	森　達哉
	早稲田大学教授	博士(情報科学)東北大学	橋本和夫