様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ５ 月 ２６ 日現在

機関番号：１１１０１
研究種目：基盤研究(C)（一般）
研究期間：2020～2022
課題番号：２０Ｋ１１８０５

研究課題名（和文）ハードウェア改竄に対して運用時のセキュリティを確保する多階層連携手法の確立と実証

研究課題名（英文）Establishment and Implementation of Multilayered Collaborative Methods to Ensure Security during Operation against Hardware Trojan

研究代表者
今井　雅（Imai, Masashi）

弘前大学・理工学研究科・教授

研究者番号：７０３２３６６５

交付決定額（研究期間全体）：（直接経費）　　3,300,000円

研究成果の概要（和文）：本研究では、正常動作時と異常動作時の各種情報を深層学習することで得られる異常動作判定器を実現するため、学習用データ自動蓄積システムの構築と、異常動作を実現するハードウェアトロイチップの設計・製造を行った。その結果、微小な差異を区別することは困難であること、限定的なハードウェアトロイにしか対応できないことが明らかになった。そこで、ハードウェアトロイによりもたらされる誤りを検知することで、障害が生じないようにする方式を提案した。蓄積した各種情報を用いた評価により、外れ値検知アルゴリズムとして1クラスSVMが有効であること、アルゴリズムの拡張により検知精度を高められる可能性があることを明らかにした。

研究成果の学術的意義や社会的意義
計算機システム構築に設計者以外の第三者が関わらざるを得なくなっており、ハードウェアに仕込まれるトロイが脅威となっている。本研究は、運用時に得られる計算機システムの各種診断情報等を用い、ハードウェアトロイが仕込まれていたとしても、正しくサービスを提供し続けられる計算機システムの実現を目的としている。研究を通し、深層学習による判定器ではハードウェアトロイの検知は限定されたものであること、ハードウェアトロイによりもたらされる誤りを外れ値検知アルゴリズムに基づいて検出する方が有効であることを明らかにした。本研究成果は様々な計算機システムに適用することができ、安全安心な情報化社会の実現に資する。

研究成果の概要（英文）：In this research, we have constructed an automatic learning data generation system and designed a hardware Trojan chip in order to realize an abnormal behavior detector obtained by a deep learning scheme in which several information under both the normal and abnormal situations are used as learning data. As a result, it became clear that it is difficult to distinguish subtle differences and that it can only be applied to the limited hardware Trojans. Therefore, we have proposed a method to detect errors caused by hardware Trojans to prevent malfunctions. Through evaluation using the obtained specific information, we have demonstrated that the one-class SVM is effective among several outlier detection algorithms. It has also been recognized that the detection accuracy can be improved by extending the algorithm.

研究分野：計算機工学

キーワード：ハードウェアトロイ　セキュリティ　ディペンダブルコンピューティング　運用時診断情報　誤り検知　外れ値検知アルゴリズム　1クラスSVM

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

コストや時間に関して厳しい制約が課される計算機システムの構築では、設計段階や製造段階において第三者の関与を完全に排除することが難しくなっており、設計者の意図しない機能を持ったハードウェア部品、ソフトウェアが情報セキュリティの脅威となっている。ソフトウェアのバグやマルウェア・ランサムウェアなど、ソフトウェア面での脅威は従来から認識されており、パターンマッチングでのウィルス検出や深層学習によるソフトウェアバグ・ソフトウェアトロイの検出など、様々な研究・対策が行われている。一方、ハードウェア面では、バックドアが仕込まれたLSIチップによる計算機システムの停止が2008年に報告されて以来、ハードウェアに仕込まれるトロイ（ハードウェアトロイ）が計算機システムにおける現実的な脅威として認識され、対策に関する研究が近年活発化している。

ハードウェアセキュリティに関するこれまでの研究のほとんどは、設計時・製造時にハードウェアトロイをいかに検出して排除するか、あるいは、PUF（Physical Unclonable Function）と呼ばれる物理的にコピーできない情報などを用いて、いかに改竄ができないようにするかという“フォールトアボイダンス（故障回避)”の概念に基づいたものとなっている。これらの手法により設計・製造時にハードウェアトロイを完全に排除することができればよいが、第三者の関与が完全に排除できない以上、ハードウェアに潜伏し続けるトロイが残る。そのため、システム運用時に構成要素のいずれかに異常出力が生じることは避けられない。しかも、時間経過に伴ってその脅威度は増加する。

従って、ハードウェアトロイがあることを前提とした計算機システムの信頼度を高めるためには、システムの構成要素にフォールト（故障）が生じても、システム全体としては正しくサービスを提供し続けることができる“フォールトトレランス（耐故障)”の概念に基づいたシステム実現が必要不可欠である。しかしながら、システム運用時のハードウェアトロイの検出手法と、活性化したハードウェアトロイの影響を抑えつつ、システムを再構築しサービスを正常に提供し続ける手法は明らかになっていない。また、ハードウェアトロイは攻撃者の意図に従って活性化されるため、物理現象を対象としていた従来のフォールトモデルとは異なるフォールトモデルに基づいた、これまでにないハードウェア・ソフトウェア協調による計算機システムが必要と考えられるが、その構成が明らかになっていない。

２．研究の目的

本研究は、計算機システムの運用時に、悪意の有無に関わらずハードウェア部品のいずれかが設計者の意図しない動作を行ったとしても、システム全体としては正しくサービスを提供し続けることができる、ハードウェア改竄に対し頑健な計算機システムの実現を目的とする。人命に関わるシステムや金融システムなどにおいては、システムの停止は１秒も認められないことがあり、企業や公共機関の情報システムなどにおいては、不正動作による秘匿情報の漏洩が膨大な損失を生むことがある。そのため、設計時ではなく、運用時に生じるハードウェアトロイの影響を皆無にする必要がある。

フォールトトレラント技術の根幹は空間的・時間的・情報的冗長化であるが、本研究ではその中でも空間的冗長性に焦点を絞る。時間的冗長化ではリアルタイム制約を満たすのが難しくなり、情報的冗長化は暗号化などが対象となるからである。ただし、単純に空間的冗長化を利用しようとすると膨大なコストがかかるため、ハードウェアトロイを適切にモデル化して評価指標を明らかにし、コストと信頼性のトレードオフを考慮した計算機システムの設計方式を提案する。本研究を通してフォールトモデルと評価指標が明らかになれば、今までは動いて当たり前だった計算機システムに対して、メーカは信頼度に応じた価格設定を行う事ができ、使用者も要求する信頼度及び必要なコストに応じた計算機システムを選択することができるようになる。従来の計算機システム市場では信頼性が価格に反映されづらくなっていたことに対し、本研究により得られるモデルと指標により、「信頼度」を価格に反映させた新たな市場価値の実現と、妥当なコストを考慮した安全安心社会の実装基盤を整えることができる。

３．研究の方法

現在のVLSIでは、各種診断情報をソフトウェア(OS含む)から確認出来るようになっている。

また、ソフトウェア・アプリケーション毎に各種ログ情報も出力されている。これら診断情報・ログ情報を用いて、ハードウェアトロイの発現をいち早く検出し、対応する方式を確立する。診断情報・ログ情報は量が膨大となるため、深層学習を用いた判定器（判定ソフトウェア）を実装する。判定器の学習には適切な教師データを必要とするため、これまでの研究で作成したハードウェアトロイを用いて、トロイ発現による影響を評価して学習データとする。また、汎用的なハードウェアトロイのモデル構築を行う。

ソフトウェアのログに関しては、例えば、Web サーバのログに関して、通常アクセスと不正アクセスの時間間隔を比較すると、不正アクセスは短い間隔に集中することが確認されている。アクセス時間間隔や接続元情報など、ソフトウェア毎に教師データとすべき情報を明らかにし、ハードウェアを含むトロイの判定を行う判定器を深層学習により実現する。

４．研究成果

(1) 機械学習を用いた非同期式回路に対するハードウェアトロイ検出について

VLSI の応用が普及する中で課題となる低消費電力化の解決策として、グローバルクロックを用いず、要求-応答ハンドシェイクプロトコルに基づいて動作する非同期式回路を利用することが一つの解として考えられる。しかし、既存のハードウェアトロイに関する研究の多くは同期式回路を対象としており、非同期式回路を対象とした研究は少なく、非同期式回路に対するハードウェアトロイの検出手法を確立することが必要となっている。

同期式回路における深層学習によるハードウェアトロイ検出手法[1]と同様、まず、ゲートレベルのネットリストの各ネットに特徴値を 51 個設定し、この特徴値を分類器の機械学習のアルゴリズムによって学習させ、未知のネットリストをノーマルネットかトロイネットかを分類する。分類器にはランダムフォレストを用いる。算出した 51 個の特徴値の重要度を基に、特に重要度の高い 10 個～15 個の特徴値を学習データとして利用した。その結果、非同期式回路に対するハードウェアトロイを検出するには、表 1 に示す 13 個を利用した検出器が最も精度がよくなることが明らかになった。この時の平均 TPR（True Positive Rate）は 76%、TNR(True Negative Rate)は 95%である。

表 1：非同期式回路に対するハードウェアトロイ検知における重要度が高い 13 個の特徴値

No.	トロイネット特徴値	説明
1	Out_loop_2	ネット n から出力方向で 1 段目のゲートが再び現れるまで 2 段かかるループの総数
2	Out_const_4	ネット n から 4 段後ろまでの固定値の総数
3	Fan_in_2	ネット n から入力側で 2 段までのロジックゲートのファンインの総数
4	In_loop_4	ネット n から入力方向で 1 段目のゲートが再び現れるまで 4 段かかるループの総数
5	In_nearest_pin	ネット n から入力側で最も近いプライマリ入力までの段数
6	Out_flipflop_3	ネット n から 3 段後ろまでのフリップフロップの総数
7	In_multiplexer_2	ネット n から 2 段手前までのマルチプレクサの総数
8	Out_Multiplexer_2	ネット n から 2 段後ろまでのマルチプレクサの総数
9	In_flipflop_1	ネット n から 1 段手前までのフリップフロップの総数
10	In_loop_2	ネット n から入力方向で 1 段目のゲートが再び現れるまで 2 段かかるループの総数
11	Out_const_2	ネット n から 2 段後ろまでの固定値の総数
12	Out_flipflop_4	ネット n から 4 段後ろまでのフリップフロップの総数
13	In_nearest_flipflop	ネット n から入力側で最も近いフリップフロップまでの段数

(2) ハードウェアトロイ挿入 VLSI チップの設計と評価について

ディスクリート半導体の真正・模造（偽造）判定には、リバースエンジニアリングを行って得られる SPICE レベルのネットリストからの判定が一つの方法として考えられる一方、著作権法の改正によりセキュリティ確保を目的としたリバースエンジニアリングは合法と認められたものの、市販部品のリバースエンジニアリングを実際に行うのは対応する企業がほとんどないことが明らかになった。そこで、TRUST-Hub ホームページに登録されているハードウェアトロイベ

ンチマークの中で、小規模回路である Micro UART（RS232）を対象として、ハードウェアトロイを挿入した回路と正常な回路の両方を、東京大学大学院工学系研究科附属システムデザイン研究センター（旧 VDEC）のチップ試作サービスを利用して、Rohm 0.18um プロセステクノロジにより試作した（図 1）。また、トランジスタレベルネットリストを用いたハードウェアトロイ検出の評価環境を構築した。

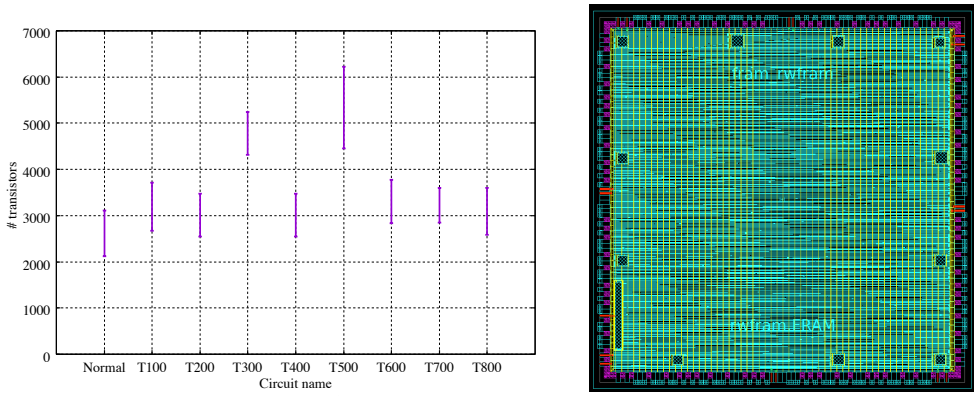

図 1：ハードウェアトロイ挿入 UART 回路のトランジスタ数とチップレイアウト

図 1 左において、Normal はトロイの挿入されていない回路、T100～T800 は TRUST-HUB ホームページに登録されているハードウェアトロイベンチマーク回路でのトランジスタ数の分布を示したものである。論理合成時のタイミング制約によって、図 1 左に示す回路規模のものが得られることが明らかになった。T300、T500 に関しては、Normal に比べて明らかにトランジスタ数が多いため、リバースエンジニアリングによりトランジスタ数をカウントすることができればT300、T500 タイプのハードウェアトロイが挿入されているであろうことが検知できる。一方、それ以外のハードウェアトロイに関しては、サイズからハードウェアトロイを検知することは困難であると言える。図 1 右は試作したハードウェアトロイ入り VLSI チップの概要を示したものである。

(3) 運用時に得られる情報に基づいた異常検知手法について

ハードウェアとソフトウェアの協調により得られる各種ログ情報を活用した異常検知に関して、単純な閾値の設定では正常か異常かの判定は難しく、大量の診断情報を用いた深層学習による判定器が一つの解と考えられるため、深層学習用のデータを自動取得する仕組みと、異常動作のモデルを構築した。

構築したシステムを用いて得られたデータを用いてハードウェアトロイの検知を試みたところ、これまでに行ってきた教師あり学習では、ハードウェアトロイの挿入されている計算機と挿入されていない計算機が必要であり、それらを用意することは非常に難しいこと、および構成が異なるものは多数存在しており、別種のハードウェアトロイでは検知できない可能性があることが確認された。

そこで、ハードウェアトロイを挿入した計算機に生じる誤りに着目し、その誤りを外れ値検知アルゴリズムにより検知する手法を検討した。想定される誤りとして消費電力増大、性能低下、外部への通信数異常を仮定し、5 種類の学習アルゴリズム、異常検知アルゴリズムを用いて評価を行った。

その結果、Isolation Forest は正常なデータから大きく逸脱した大きな誤りの検知の精度が高く、主成分分析は想定した誤り全てで精度が非常に低くなった。そして、高速最小共分散行列式は通信数異常の検知において Isolation Forest と似たような特徴を持ち、大きな誤りの検知の精度が高く、局所外れ値因子法と 1 クラス SVM はどちらも正常なデータに対しての検知精度が他のアルゴリズムと比較して低めであるものの、小さな誤りに対してでも安定して高い精度で検知を行えることがわかった。これらのアルゴリズムは、想定する誤りの種類やその大きさによって検知に最適なアルゴリズムが変わるが、これらの中でも特に 1 クラス SVM は、消費電力増大、通信数異常、性能低下の全てを安定して高い精度で検知することが可能であり、誤りが小さなものでも精度が低くなりにくいことがわかった。そのため、本研究で想定した誤りの検知には、誤りの大きさが不明である場合は 1 クラス SVM が最も適していると言える（図 2）。しかし、1 クラス SVM は正常なデータに対しての精度が他のアルゴリズムと比較して低いという欠点を持

つ。この欠点の対策としては、他のアルゴリズムでの判定結果を組み合わせることで正常なデータに対する精度を補うという手法が挙げられる。

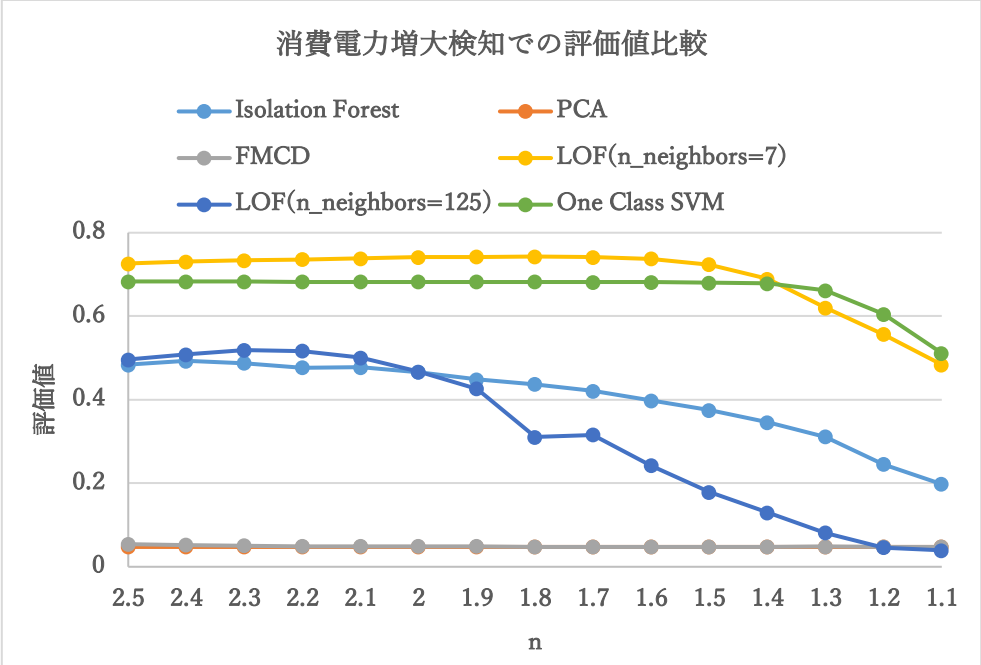

図 2：消費電力増大検知でのアルゴリズム別検知率

＜参考文献＞

[1] Kento Hasegawa, Masao Yanagisawa, and Nozomu Togawa. “Trojan-net Feature Extraction and its Application to Hardware-Trojan Detection for Gate-level Netlists using Random Forest.” *IEICE Trans. Fundamentals*, E100-A(12):2857-2867, Dec. 2017.

５．主な発表論文等

〔雑誌論文〕　計0件

〔学会発表〕　計4件（うち招待講演　0件／うち国際学会　3件）

１．発表者名 Masashi Imai
２．発表標題 Asynchronous Building Blocks and Their Application for Ultra Low Energy Devices
３．学会等名 NOLTA2020（国際学会）
４．発表年 2020年

１．発表者名 Masashi Imai
２．発表標題 Novel Circuit Structure of Basic Standard Cells against Glitches
３．学会等名 ASYNC2021（国際学会）
４．発表年 2021年

１．発表者名 Takuro Kasai, Masashi Imai
２．発表標題 Development of Diagnosis-based Hardware Trojan Tolerate System
３．学会等名 SASIMI2022（国際学会）
４．発表年 2022年

１．発表者名 葛西巧朗，今井雅
２．発表標題 外れ値検知アルゴリズムによるハードウェアトロイ内包計算機に生じる誤りの検知
３．学会等名 情報処理学会東北支部研究報告
４．発表年 2023年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関