May-June 2020
ISSN: 0193-4120 Page No. 9702 – 9710

Trojan Detection System

Muneeswaran V -1Associate Professor, Sri Krishna College of Engineering and Technology
Department of Computer Science and Engineering
Dr.K.Sasi Kala Rani- 2 Professor, Sri Krishna College of Engineering and Technology
Department of Computer Science and Engineering
Dhanush Vijay S P- 3 UG Student Sri Krishna College of Engineering and Technology.
Department of Computer Science and Engineering
Saran M- 4 UG Student Sri Krishna College of Engineering and Technology
Department of Computer Science and Engineering
Ajay Kumar S - 5 UG Student Sri Krishna College of Engineering and Technology
Department of Computer Science and Engineering

Article Info
Volume 83
Page Number: 9702 – 9710
Publication Issue:
May - June 2020

Abstract:
Hardware security has become an important concern in recent years. A hardware Trojan (HT) is also a hardware virus. It is a malicious modification of a circuit so it can manage, modify, disable, monitor or affect the operation of the circuit. HTs can be inserted into IC during RTL or during manufacturing, through manipulation of the layout masks and varying the doping concentration. As adversaries would want access to foundries to insert Trojans during the fabrication process, the likelihood of them being inserted at design time is far higher. it is vital to detect the hardware Trojans from a viewpoint of security. Due to the outsourcing in hardware production, malicious circuits (or hardware Trojans) are easily inserted into hardware products by attackers. Since hardware Trojans detection is difficult. Under the circumstances, numerous hardware-Trojan detection methods are proposed. During this project, it elaborates the overview of hardware-Trojan detection and reviews the hardware-Trojan detection methods using machine learning (both supervised and unsupervised learning). In supervised learning, applying KNN algorithms to resolve this problem. In unsupervised learning, applying neural networks to resolve this problem. a way to detect Hardware Trojan with the help of Deep Learning algorithm by feeding it features extracted from the gate-level netlist of the circuit. The proposed method doesn't require any golden circuits (reference circuits) circuits. The features that are extracted from the gate-level netlist are accustomed to train the Deep learning algorithm. This method doesn't require the simulation of the circuit so on classify genuine nodes and Trojan affected nodes.

Article History
Article Received*: 19 November 2019*
Revised*: 27 January 2020*
Accepted*: 24 February 2020*
Publication*: 18 May 2020*

I. INTRODUCTION

A Hardware Trojan could be a malicious modification of the circuitry of an integrated circuit. A hardware Trojan is totally characterized by its physical representation and its behavior. The payload of an HT is that the entire activity that the Trojan executes when it's triggered. Hardware Trojans could also be introduced as hidden "Front-doors" that are unknowingly inserted while designing a computer chip, by employing a pre-made application-specific computer circuit (ASIC) semiconductor intellectual property core (IP Core) that are purchased from a non-reputable source, or inserted internally by a rogue employee, either performing on their own, or on behalf of rogue interest groups, or state sponsored spying and espionage.The threat of a heavy, malicious, design alteration is especially relevant to government agencies. Resolving doubt about hardware integrity is one way to scale back technology vulnerabilities

May-June 2020
ISSN: 0193-4120 Page No. 9702 – 9710

within the military, finance, energy and political sectors of an economy. Since fabrication of integrated circuits in untrustworthy factories is common, advanced detection techniques have emerged to get when an adversary has hidden additional components in, or otherwise sabotaged, the circuit's function.

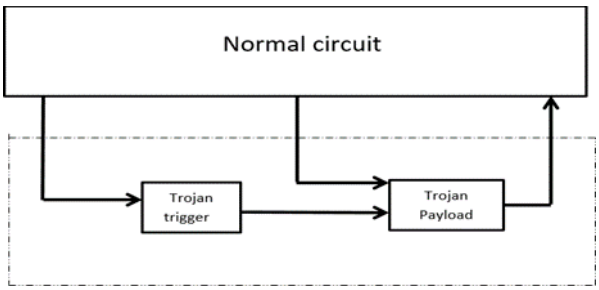

Fig 1 Trojan Infected Circuit

Hardware Trojans are a great threat to any hardware. It is just like a virus in a software which crashes the computer and steals secret data from the computer. These third parties who inject Hardware Trojan into the hardware are called adversaries. As many detection methods have come to detect Hardware Trojans from hardware, the adversary also has become more intelligent and inserting the Hardware Trojans which triggers very rarely.

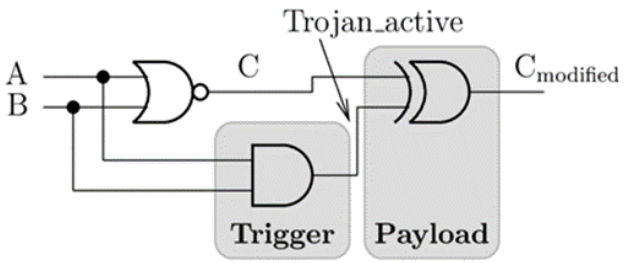

Fig 2 Hardware Trojan

NETLIST

In electronic design, a netlist could be a description of the connectivity of an electronic circuit. In its simplest form, a netlist consists of a listing of the electronic components in an exceedingly circuit and a listing of the nodes they're connected. A network could be a collection of two or more interconnected components. The structure and its representation of netlists can vary considerably, but the elemental purpose of each netlist is to convey connectivity information. Netlists usually provide nothing but quiet instances, nodes, and maybe some attributes of the components involved. If they express way more than this, they're usually considered to be a hardware description language like Verilog, or one in every of several languages specifically designed for input to simulators.

VERILOG

Verilog, standardized as IEEE 1364, may be a hardware description language accustomed to model electronic systems. Hardware description languages like Verilog are kind of like software programming languages because they include ways of describing the propagation time and signal strengths (sensitivity). There are two styles of assignment operators; a blocking (=), and a non-blocking (<=) assignment. The non-blocking assignment allows designers to explain a state-machine update while not having to declare and use temporary storage variables. Since these concepts are a part of Verilog's language semantics, designers could quickly write descriptions of enormous circuits during a relatively compact and concise form. At the time of Verilog's introduction (1984), Verilog represented an amazing productivity improvement for circuit designers who were already using graphical schematic capture software and specially written software programs to document and simulate electronic circuits.

NEURAL NETWORK

A neural network is similar to the human brain's neural network. It could be a mathematical relation that collects and classifies information in line with a particular architecture. The network bears a robust resemblance to statistical methods like curve fitting and multivariate analysis. It could be a series of algorithms that endeavors to acknowledge underlying relationships during a set of information through a process that mimics the way the human

May-June 2020
ISSN: 0193-4120 Page No. 9702 – 9710

brain operates. During this sense, neural networks see systems of neurons, either organic or artificial in nature. Neural networks can adapt to changing input; so the network generates the simplest possible result without having to revamp the output criteria. The concept of neural networks, which has its roots in computer science, is swiftly gaining popularity within the development of trading systems.

II. LITERATURE SURVEY

PHYSICAL CHARACTERISTICS

Physical characteristics are nothing but a physical representation of a Trojan. Hence, Hardware Trojans are further classified into Distribution, size and type depend on its physical appearance. Sometimes, Trojan circuits might be big enough to be visible. So, the Trojan designer scatters the Trojan components across the IC disabling various functions. This type of Trojan is called loosely distributed Trojan. On some occasions, the Trojan circuit is so small that its components occupy the chip's layout. This type of Trojan is called tightly distributed Trojan.

1. Size: Hardware Trojan's size can be manipulated by the number of circuit components it has. A Trojan may be small, medium or big.

2. Structure: The injections of Hardware Trojans into a hardware is the hectic job even for the well trained adversary. So, If the adversary is not in the mood, he/she might even alter the dimension of the chip by injecting Hardware Trojans. Although it rarely happens as it is easy to detect hardware Trojans if the chip's dimension had been changed. This type of Trojan is called structural Trojan.**3. Type:** A Trojan can be classified into Functional Trojan or Parametric Trojan based on its type. If the original chip is modified by adding or deleting gates or transistors, then it is called Functional Trojan. The Parametric Trojan, changes the first hardware, e.g. diminishing of wires, debilitating of Flip-Flops or transistors or utilizing Focused Ion-Beams (FIB) to lessen the unwavering quality of a chip.

ACTIVATION CHARACTERISTICS

1. Trigger: Trojan trigger is a component of a HT. Trojan trigger activates the payload of the HT. The trigger rarely triggers so that it will be difficult for us to detect the Trojan part in the hardware. Trojan payload can be triggered both externally and internally. Internal trigger activates the payload when a few of the low toggle nodes in the IC toggles. External triggers are nothing but external signals from outside of the hardware which activates the payload. Example of an external trigger is RF signal.

2. Behaviour: Behavioural Trojans are classified into always on, condition based, time delay. An always on Trojan can be a lessened wire. A chip that is adjusted thusly creates blunders or flops each time the wire is utilized strongly. In the case of time delay Trojan, the Trojan trigger triggers the payload after a counter logic. So, it takes several time spans, after which the Trojan payload gets activated. A conditional based Trojans are activated by internal or external triggers based on a condition which can be a counter, toggling etc.,

ACTION CHARACTERISTICS

1. Function Modification: In the function modification category, the function of the IC is modified because of the Trojan injection.

2. Specification Modification: Specification modification, the chip's parametric properties can be changed, when the Trojan payload is activated. For example: A process delay can be incited.

3. Data Transmission: In this category, a secret information, key or data stored or used in your hardware can be transmitted to the adversary in the form of a signal (eg. RF signal).

FUNCTIONAL TESTING

May-June 2020
ISSN: 0193-4120 Page No. 9702 – 9710

1. Automatic Test Pattern Generation (ATPG): This method detects Trojan by checking the logic pattern of the output while giving all possible inputs to the input port of IC. An input vector is given as an input to the chip. The corresponding output is inspected to detect the faults in a circuit. This is applied to most test cases. Since it detects Trojans by checking the operation of the circuit, the proposed technique can't detect functional Trojans as the circuit is functional in all test cases. This technique cannot find the particular input vector which activates the Trojan trigger, because all the test cases can't be tested as it simply takes too much time. This is one of the widely used detection techniques in the fabricated ICs.

2. Built-In-Self-Test Techniques: This method is like preserving the chip even before it is made. The Built-in circuit is employed in the original chip design to scan the unwanted signals emitted from the chip. This additional circuit to protect the chip finds the manufacturing faults. It can also detect malicious circuits in the IC while testing. Even Design hardware trust does the same, but complex circuits are added to the original chip design. There are different methods available.

3. Side Channel Analysis: An integrated circuit consists of all components which emit electromagnetic signals. When IC is supplied with electric current, it emits these electromagnetic signals. The adversary manipulates and analyses these signals to get info about the data which the IC processes. Power consumption, Timing information, Electromagnetic signals and heat analysis are examples of side channel signals. The functional Trojans can be detected by the calculation of these side channel signals emitting from the IC. The values found from calculation are utilized as a signature for the device which is already analysed.

4. Genuine Chips: The detection method involves comparison of values of specific parameters between the original IC and the test ICs. The original ICs are assumed to be Trojan free ICs. The original ICs and test ICs should have been manufactured in the same fabrication factory. Because the original IC produced from a factory compared to test ICs produced from different factories have variances.

5. Path Delay Fingerprinting: In this method fingerprint is generated from path delay between the pins in the chip. In another method, the fingerprint can also be generated by measuring the path delay between two registers in a chip. The timing information is utilized to measure the path delay between the registers. Theoretically, the timing behaviour of the IC can be analysed by utilizing the netlist.

III. EXISTING SYSTEM

Hardware devices are susceptible to the installation of trojans during the manufacturing phase. Trojans affect the hardware by creating back doors that compromise the security of the device or by overheating the hardware component and decreasing its efficiency. Outsourcing the manufacturing process of hardware ensures greater risk of trojan installation to the organisation. There are multiple types of trojans whose detection can be crucial in the success of the manufactured hardware devices.

The majority of hardware components that are manufactured are being outsourced to other organisations to improve the efficiency of the product manufacturing process. There is an increased possibility of trojans being installed to the hardware components during the manufacturing process. Trojans can affect the performance of the devices by causing heating problems, some trojans create security issues by accessing a back-door. Trojans can also decrease the overall lifespan of the hardware devices and result in losses in revenue to the companies.

May-June 2020
ISSN: 0193-4120 Page No. 9702 – 9710

DISADVANTAGES

- Trojans can create security issues in the devices.
- Hardware can become overheated due to the presence of trojans.
- The lifespan of the hardware is reduced.

IV. PROPOSED SYSTEM

Detecting Hardware Trojans at an RTL design phase. So, Verilog files of many circuits (both Trojan free and Trojan affected) in TRUST HUB. Converted these Verilog files (.v) to Gate-level netlists (.v) through Synopsys Design Vision tool. A Verilog file is a behavioural model of a circuit whereas the netlist is a gate-level description of a circuit. These netlists are given to a feature extraction algorithm to extract features from these files. Feature extraction algorithm takes each netlist (.v) and another file (.yaml) as an input and extracts features for each net of each netlist and stores the feature values for each net of each netlist respectively as (.csv) files. An example of a (.csv) file in table (4.1) where the 0th column indicates each net and 1st column to 8th column indicates values of each feature for each net. Given this (.csv) file as an input to a Deep learning algorithm. So, first decide few of the (.csv) files which can be used for training and the remaining which can be used for testing.

1	2	3	4	5	6	7	8	9
U7	0.55	0.47	0.38	0.69	0.44	0.44	7.14	7.0
U8	0.66	0.54	0.41	0.75	0.51	0.51	8.96	7.5
U9	0.44	0.43	0.36	0.64	0.26	0.26	6.12	6.5

Table 1 Converted File

If type '0' in a particular row in that additional column, it means that particular net is not affected with Trojan and if type '1', it means that the particular net is a Trojan affected net as shown in table 2.

1	2	3	4	5	6	7	8	9
0.55	0.47	0.38	0.69	0.44	0.44	7.14	7.0	0
0.66	0.54	0.41	0.75	0.51	0.51	8.96	7.5	0
0.44	0.43	0.36	0.64	0.26	0.26	6.12	6.5	0

Table 2 Test File

From table 2 the user can see that all the nets U7, U8, U9 are Trojan free nets. So, First gave this (.csv) file as an input to the algorithm for training. In the testing phase. Given (.csv) files which were selected for testing as shown in table 3 as an input to the algorithm.

0	1	2	3	4	5	6	7	8
U74	0.005	-9.22	0.071	0	0	1.543	0.6923	1
U75	0.0053	-9.83E	0.0711	0	0	1.5430	0.6923	1
U76	0.0053	1.49E	0.0711	0	0	1.5430	0.6923	1
U77	0.0053	6.14E	0.0711	0	0	1.5430	0.6923	1
U78	0.005	3.59E	0.071	0	0	1.543	0.6923	1
U79	0.005	4.58E	0.071	0	0	1.543	0.6923	1

May-June 2020
ISSN: 0193-4120 Page No. 9702 – 9710

U80	0.005	7.53E	0.071	0	0	1.543	0.6923	1
U81	0.005	7.14E	0.071	0	0	1.543	0.6923	1

Table 3 Trojan File

From table 3 the user can see that the 0th column indicates each net and the 1st to 8th column indicates features of each net. Since it is training, don't need to give the classification as the algorithm will classify each net. Output of an algorithm is shown in table 4.

0	1	2
1	U74	1
2	U75	1
3	U76	1
4	U77	1
5	U78	1
6	U79	1
7	U80	1
8	U81	1

Table 4 Result File

As the user can see from the table 4, the 1st column indicates each net and the 2nd column indicates the Trojan classification. From the output can clearly see that the algorithm classified each and every net as a Trojan affected net. But, that's not possible because the third-party vendor inserts very rarely triggered Trojans which don't affect all the nets of a circuit. For now, use only one netlist for training and one for testing. For the algorithm to classify correctly, have to train the algorithm with more circuits.

GRAPH THEORY

Graph theory is the study of graphs mathematically. A graph consists of vertices and edges. The vertices are connected by the edges to form a graph. There are many types of graphs like planar graph, line graph, directed graph etc., Netlist consists of components like resistors, flip flops, gates etc., and wires. components are all connected with wires. Hence, consider the components as nets and wires as edges to form a graph. Extract the relevant features from the graph.

ANACONDA DISTRIBUTION

It is an open-source distribution of the Python and R programming languages for scientific computing, that aims to simplify package management and deployment. Package versions are managed by the package management system conda. it's no IDE of its own. The default IDE bundled with Anaconda is Spyder which is solely another Python package that will be installed even without Anaconda. Anaconda distribution comes with 1,500 packages selected from PyPI additionally because the conda package and virtual environment manager. It also includes a GUI, Anaconda Navigator The big difference between conda and the pip package manager is in how package dependencies are managed, which can be a big challenge for Python data science and so the explanation conda exists.

CONFUSION MATRIX

A confusion matrix may be a table that's often accustomed to describe the performance of a classification model (or "classifier") on a collection of test data that verity values are known. In the field of machine learning and specifically the matter of statistical classification, a confusion matrix, also called a slip matrix, may be a specific table layout that permits visualization of the performance of an algorithm, typically a supervised learning one. Confusion matrix consists of a table with two rows and two columns that reports the quantity of false positives, false negatives, true positives, and true negatives. This permits more detailed analysis than mere proportion of correct classifications (accuracy). Accuracy will yield misleading results if the information set is unbalanced; that's, when the

May-June 2020
ISSN: 0193-4120 Page No. 9702 – 9710

numbers of observations in several classes vary greatly. For instance, if there have been 95 cats and only 5 dogs within the data, a selected classifier might classify all the observations as cats. the accuracy would be 95%, but in additional detail the classifier would have a 100% recognition rate (sensitivity) for the cat but 0% recognition rate for the dog class.

DEEP LEARNING

Deep learning comes under the topic of machine learning which comes under Artificial Intelligence. Deep learning simulates the process of a human brain. It is an artificial neural network. It is mainly used for classification of samples. Deep learning, which is a part of machine learning algorithms, is also called Deep Neural Network. Data science is the next big sector that is rising and to be boomed in the next coming decades. Big data, the whole lot of data available in the internet, social media, e-commerce etc., can be used to retrieve useful information. This big data can be shared via cloud computing but the data is so unstructured that humans cannot understand and so can't retrieve useful information which they want from it. The leading organizations in the world are participating in this challenge of analysing this big data and they are developing their own Artificial Intelligence for almost every field in the industry. Example is Google's Deep mind.

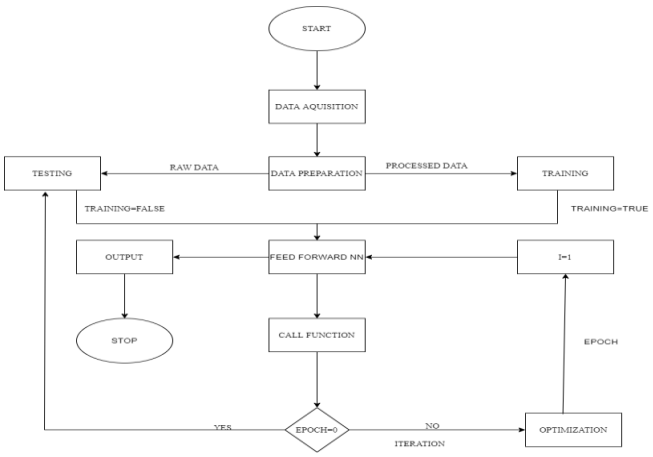

Fig 3 Deep Learning Algorithm.

IMPORTANCE OF PROPOSED SYSTEM

Since the 1990s, there has been a steady trend in outsourcing various aspects of design, fabrication, testing and packaging of IC instead of in-house IC designs. This brings unknown security threats and trust concerns in the ICs. Hardware Trojans have become a major threat faced by most VLSI designers. These Trojans are designed to act as silicon time bombs to disable IC, Intellectual property and IC piracy, untrustworthy third-party ICs. They are designed to be triggered by very rare logic conditions at internal nodes of the circuit. Hence there is a need for techniques and algorithms for detection of rarely triggered Trojans. Soft computing techniques are widely used in malware detection these days. These techniques have the ability of learning from past instances and can categorize normal and abnormal behaviour. In this project, focus on the testing for Trojans by using Deep learning techniques. In most cases the golden chip is not available and this technique doesn't require the golden chip hence is a very huge advantage.

ADVANTAGES

- The system can be used to detect trojans at the manufacturing stage.
- The losses incurred by trojans are removed.
- It can improve the efficiency of the hardware devices.

V. ANALYSIS AND RESULT

The number of hidden layers in the middle and the nodes present in it greatly affects the decision making of the deep neural network. It is clear that the quantity of hidden layers and hidden nodes makes the algorithm to understand the data more, but it also makes the neural network to over fit the data. Even though this overfitting is largely restricted by the dropout layer, it takes too much time for the neural network to process. Other than this, there is also a great chance of gradient problem [1]. So,

May-June 2020
ISSN: 0193-4120 Page No. 9702 – 9710

decide the optimum quantity of hidden layers and hidden nodes.

SECTION I

Firstly, tried to train the benchmark circuits with two hidden layers. Each node has values (16, 32), (32, 16), (32, 64), (64, 32), (128, 64), (64, 128), (256, 128), (128, 256). Where the first value in the braces represents the number of hidden nodes of the 1st hidden layer and the second value represents the number of hidden nodes of the 2nd hidden layer. The results are given below in table 5.

No. of nodes in 1st layer	No of nodes in 2nd layer	True positive rate (%)	True negative rate (%)
16	32	76	98
32	16	96	89
32	64	85	97
64	32	97	85
128	64	98	87
64	128	94	90
128	256	93	90
256	128	99	83

Table 5 Results of neural network with two hidden layers

Secondly, train the benchmark circuits with three hidden layers. Each node has values (16, 32, 16), (32, 64, 32), (64, 128, 64), (128, 256, 128), (256, 512, 256). Where first value in the braces represents number of hidden nodes of 1st hidden layer and second value represent number of hidden nodes of 2nd hidden layer and second value represents number of hidden nodes of 3rd hidden layer. The results are given below in table 6.

No of nodes in 1st layer	No of nodes in 2nd layer	No of nodes in 3rd layer	True positive rate (%)	True negative rate (%)
16	32	16	88	96
32	64	32	90	92
64	128	64	99	76
128	256	128	97	93
256	512	256	99	84

Table 6 Results of neural network with three hidden layers

SECTION II

The netlists which used for training and its label information and the netlists which used for testing and its label information are also shown in table 7.

Netlists used for testing	Total no. of nets	No of Trojan nets
RS232-T2000	221	10
RS232-T1900	227	13
RS232-T1700	218	8

Table 7 The netlists and its label information

The same procedure is followed in section II: Trained the benchmark circuits with two hidden layers. Each node has values (16, 32), (32, 16), (32, 64), (64, 32), (128, 64), (64, 128), (256, 128),(128, 256). Where the first value in the braces represents the number of hidden nodes of the 1st hidden layer and the second value represents the number of hidden nodes of the 2nd hidden layer. The results are given below in table 8.

No. of nodes in 1st layer	No of nodes in 2nd layer	True positive rate (%)	True negative rate (%)
16	32	69	90
32	16	90	84
32	64	91	82
64	32	93	80
128	64	87	93
64	128	88	94

May-June 2020
ISSN: 0193-4120 Page No. 9702 – 9710

128	256	82	86
256	128	78	90

Table 8 Results of neural network with two hidden layers

Trained the benchmark circuits with three hidden layers. Each node has values (16, 32, 16), (32, 64, 32), (64, 128, 64), (128, 256, 128), (256, 512, 256). Where first value in the braces represents number of hidden nodes of 1st hidden layer and second value represents number of hidden nodes of 2nd hidden layer and second value represents number of hidden nodes of 3rd hidden layer. The results are given below in table 9.

No of nodes in 1st layer	No of nodes in 2nd layer	No of nodes in 3rd layer	True positive rate (%)	True negative rate (%)
16	32	16	77	84
32	64	32	85	90
64	128	64	94	88
128	256	128	92	83
256	512	256	90	86

Table 9 Results of neural network with three hidden layer

VI. CONCLUSION

The trojan detection system is used to identify the presence of harmful trojans in the hardware devices during the manufacturing stage and helps in the isolation of the devices that are affected by these trojans. The system helps in eliminating problems like overheating in the hardware and reduces the revenue lost by the presence of trojans. Detection of Hardware Trojan with the help of Deep Learning algorithm by feeding it features extracted from the gate-level netlist of the circuit first, extracted 8 features from the netlists. The features are very important in deep learning as it is the input to the neural network. The classifier decides the class of the sample depends on its features. Hence, there is a possibility that one or two features might be irrelevant or nearly the same which might confuse the deep neural network. So, the features should be optimized and hence we use dimensionality reduction to optimize the number of features fed to the neural network. We used dimensionality reduction algorithms to find the best set of features among them. We used those features to classify the Trojan nets among the normal nets using deep neural networks. The results have been better than the previous results produced by the above mentioned machine learning algorithms.

References

1. Tom Kean, David McLaren and Carol Marsh: Verifying theAuthenticity of Chip Designs with the DesignTag System
2. C. Fagot, O. Gascuel, P. Girard and C. Landrault: On Calculating Efficient LFSR Seeds for Built-In Self Test, Proc. Of European Test Workshop, 1999, pp 7-14.
3. W. T. Cheng, M. Sharma, T. Rinderknecht and C. Hill: Signature Based Diagnosis for Logic BIST, ITC 2006, Oct. 2006, pp. 1-9.
4. J.Zhang, F. Yuan, and Q. Xu, "DeTrust: defeating hardware trust verificationwithstealthyimplicitly-triggeredhardwareTrojans,"in Proc. ACM
5. SIGSAC Conference on Computer and Communications Security (ACM-CCS), pp. 153–166, 2014.
6. Kwang-Il Goh, Byungnam Kahng and Doochul Kim Universal behavior of Load Distribution in Scale-Free Networks. Physical Review Letters 87(27):1–4, 2001.
7. Mark E. J. Newman: Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Physical Review E 64, 016132,2001.
8. Boldi, Paolo, and Sebastiano Vigna. "Axioms For Centrality." Internet Mathematics 10.3-4 (2014): 222-262.