International Journal of High Speed Electronics and Systems
Vol. 31, Nos. 1–4 (2022) 2240003 (9 pages)

DOI: 10.1142/S0129156422400031

World Scientific
www.worldscientific.com

Reverse Engineering Protection Using Obfuscation Through Electromagnetic Interference

William Stark*, Shuai Chen† and Lei Wang‡

Department of Electrical and Computer Engineering, School of Engineering, University of Connecticut, Storrs, Connecticut 06269, USA
*william.stark_jr@uconn.edu
†shuai.chen@uconn.edu
‡lei.3.wang@uconn.edu

Received 31 May 2022
Accepted 31 May 2022

This paper discusses commonly used reverse engineering methods to illegally recreate printed circuit board (PCB) designs. A solution using transformative electronics is presented to prevent the discussed reverse engineering methods by obfuscating the design. The transformative electronics solution is employed in a specific application that results in a reverse engineered board to be incorrectly recreated, where the signals would be distorted due to added electromagnetic interference (EMI). The non-conductive vias that are part of the obfuscation would allow the inclusion of EMI generators that would not affect the circuit in an original design but would prevent copied designs from working correctly. A machine learning algorithm is being designed to optimize the placement of the EMI sources in an original PCB.

Keywords: PCB; obfuscate; reverse engineer; EMI; machine learning.

1. Introduction

As the electronics industry continues its never-ending expansion, electronic hardware finds increasingly numerous applications that involve embedded circuit boards. With this expansion comes increased production, designs, and an unwanted growth of industrial espionage. Design theft has become a more significant concern with products that contain printed circuit boards (PCBs), with up to 90% of companies that use PCBs in their products experiencing some level of intellectual property infringement. This level of brutal design copying is due to easy reverse engineering. PCB designs can be taken apart, scanned and recreated in circuit design software, and produced en masse, effectively stealing the work from the original designers. As a result, different solutions have been sought out to protect a PCB from being reverse engineered. This paper seeks to prevent illegal reverse engineering through electromagnetic interference using transformable electronics.

*Corresponding author.

2. Reverse Engineering Methods

There are several strategies that have been employed to reverse engineer a PCB. By simply acquiring an example of a product, one can deduce a substantial amount of information from the physical appearance of the board, the metal traces, and the component placement. However, this alone is often not enough to reverse engineer a PCB, due to the intermediate layers that contain additional trace connections between components. To examine these layers, an attacker may perform an x-ray scan of the PCB, taking pictures of each layer. This will allow them to map out every electrical connection in the board and create a complete circuit replication. By having images of each layer, advanced algorithms can be used that accurately recreate the circuit structure in a design. The board can be further probed at each node to understand its electrical characteristics and make design duplication that is more accurate to the original. This is considered a nondestructive form of reverse engineering.

Another form of reverse engineering is delayering the PCB by physically scraping off the outer layers, exposing the internal electrical connections [1], which renders the PCB partially unusable. However, the end result is the same, where pictures can be taken of each layer throughout the removal process and an algorithm can be used to generate a model of the circuit.

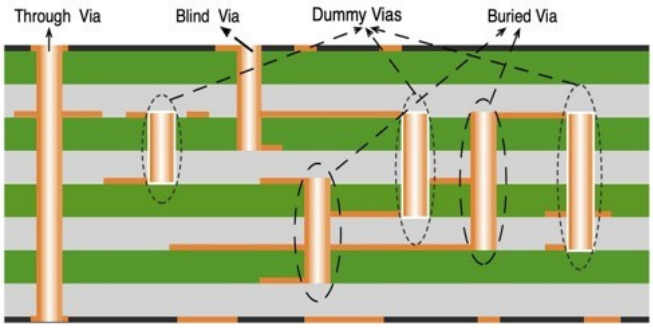

Fig. 1. Insertion of transformable and faux vias.

3. Obfuscation Strategy

Both destructive and non-destructive reverse engineering methods are extremely easy to accomplish and difficult to prevent. However, by attempting to hide the circuit design through physical obfuscation, the chance of an attacker correctly reverse engineering a PCB design can be greatly reduced. This paper seeks to take on both previously described reverse engineering techniques with a common solution. To confuse algorithms that analyze circuit design from photos of each layer, the vertical connections between PCB layers, known as vias, can be inserted throughout the layers. Instead of typical vias, these will be of a non-conductive material. These vias will not interfere with the normal function of the circuit, as they do not conduct. However, they can be placed to appear as though

they are connecting traces between layers, making it much more difficult for an algorithm to distinguish which vias complete the true circuit and which ones are dummy connections. This significantly increases the possibility of circuit combinations and would require an unrealistic amount of processing resources to compute every possible layout.

Most PCB vias use typical conductor material such as copper. These vias can be replaced with another conductive material, in this case magnesium. With a resistivity of 44.7nΩ·m [2], magnesium can provide similar conductive properties as copper; however, where it is unlike copper is its reactivity with other elements. In the event that an attacker attempts to destructively reverse engineer a PCB, the interior layers will suddenly be exposed to air, whereupon the magnesium will react to become magnesium oxide, which is not conductive. Therefore, by combining magnesium vias serving as the true connections with dummy vias made of magnesium oxide, the PCB can automatically protect itself from both forms of reverse engineering and obfuscate its true design. This solution is very easy to implement, with low overhead cost and a simple triggering mechanism that can provide vastly superior protection to a design than currently existing countermeasures [3].

4. Protection Solution Application

The usage of transformative via materials can be applied to a specific protection solution. Adding additional isolated traces in proximity to functional traces of the design can be combined with transformative vias to further obfuscate circuit operation. These additional traces would be connected to the circuit through the nonconductive MgO vias and have no effect on the functionality of an original design. However, if a PCB is recreated using images obtained through nondestructive reverse engineering, then the copied design will include these additional traces. Because the MgO vias would confuse the attackers by appearing as standard conductive vias, this copied design would recreate the circuit incorrectly, and have pairs of generator and receptor traces that would introduce electromagnetic interference (EMI) to the trace. Distortion on key traces of a circuit can hide signals that are critical to operation, or signals that may contain sensitive information in the event of a communication hack, rendering the copied design useless and preventing successful distribution.

A generic generator/receptor model (see Fig. 2) can be derived to determine the near-end voltage as

$$V_{NE}(t) = \frac{R_{NE}}{R_{NE}R_{FE}} L_{GR} \frac{dI_G}{dt} + \frac{R_{NE}R_{FE}}{R_{NE} + R_{FE}} C_{GR} \frac{dV_G}{dt} \tag{1}$$

$$V_{FE}(t) = \frac{R_{FE}}{R_{NE}R_{FE}} L_{GR} \frac{dI_G}{dt} + \frac{R_{NE}R_{FE}}{R_{NE} + R_{FE}} C_{GR} \frac{dV_G}{dt} \tag{2}$$

which result on the receiving trace. Experimentation using this solution has already been conducted using multiple original signal frequencies, with results showing that adjacent trace pairs, when configured where one will receive noise from another, can distort signals up to 43.6 percent from its original shape.

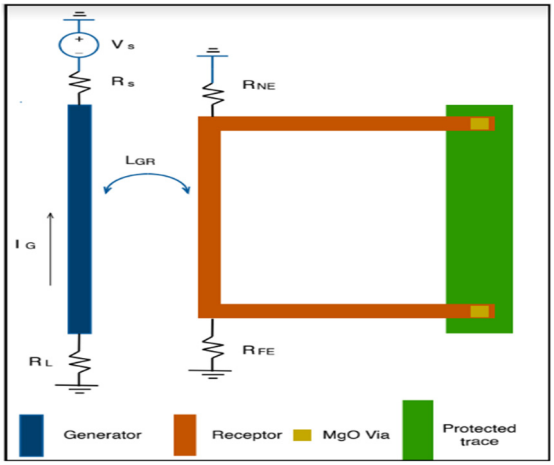

Fig. 2. Depiction of pairing a generator and receptor trace to create crosstalk.

5. Optimization of Solution Application

Generator traces can be an effective application for obfuscation using transformable electronics. The adjacent generator traces must be optimized to be at the most vulnerable circuit locations to cause the largest amount of EMI distortion possible. This optimal placement of generators can be solved using machine learning to determine parameter values. Algorithms can be designed to take in circuit properties and resulting distortion on traces, and use the data to build a model that places traces for maximum EMI.

With multiple parameters to consider, and with each of those having a varying amount of influence on the EMI, a machine learning algorithm employing multivariate polynomial regression is selected, taking the form

$$y_i = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_n x^n \tag{3}$$

where the parameter data is represented in an X matrix with the y vector being the resulting EMI distortion. Multidimensional polynomial regression can often have difficult solutions to determine and can be negatively impacted by parameter selection. To improve the curve fitting, several estimation tools will be tested, such as ridge regression due to the potential for parameters to be correlated in many circuit designs.

The significance of using a machine learning algorithm to maximize generator placement is that a variety of design parameters can be the basis for an optimal design. The objective function used in machine learning is an approximation that attempts to mathematically relate parameters that may otherwise not have a described connection in a contextualized application. By being able to rationalize a large enough data set, a general pattern for the output can be determined based on the input parameters. This can further allow one to determine which parameter settings will result in a desired output using the discovered relationship. By determining a relationship between multiple design parameters

and the resulting EMI distortion, this work alters those parameters to achieve distortion for obfuscation.

A machine learning model is built using a selection of parameters that are known to have a plausible connection to potential distortion, and are often taken into consideration when PCB designers are normally attempting to limit EMI in their circuits. These parameters include, but may in the future not be limited to, the frequency range experienced at a trace, the number of connections to the nodes of a trace and the length of the trace. Included in these parameters is an additional parameter that will be investigated, that being the length of the generator trace paired with the design. The related result of these parameters would yield a distortion level. Furthermore, if a considered trace is connected to an IC chip, which is known to be a potential source of EMI, a baseline distortion level is added to the result.

The plan for data collection requires gathering and extracting these parameters from PCB designs, with the circuits discretized as a netlist. This is made possible by SPICE software, where frequency/switching frequency analysis of circuits can be accomplished. The quantity of node connections and trace lengths can be collected from PCB board design files that specify the physical characteristics and appearance of a PCB. Additional work will be needed to calculate the trace length directly from the dimensions of each trace segment provided in the design file.

While not an ever present component throughout an entire PCB design, IC chips are a noteworthy source of interference. This is due to their switching noise and high frequency internal signals. An estimation relationship is based on [4], which describes an IC as a Norton Equivalent model (see Fig. 3) for the overall voltage, current and impedance as observed from points in the PCB. The methods provided in [4] can be supplied with values extracted from SPICE simulations of the circuit, provided an equivalent IC model is present in the library. The resulting noise source signal V_s can be calculated as a baseline addition to any distortion contributed on PCB traces for locations in the circuit directly connected to an IC pin. The presence of an IC connection at a trace will contribute to determining the optimization of EMI generation in the machine learning model.

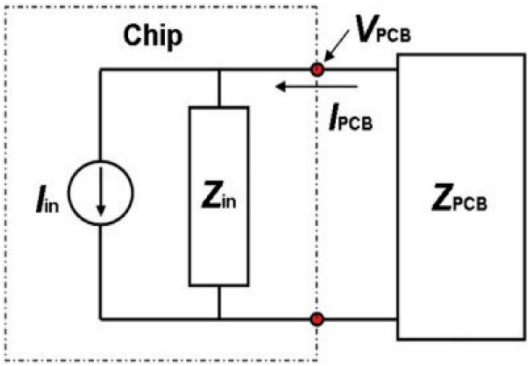

Fig. 3. Norton equivalent representation of integrated circuit voltage noise source.

Because the EMI distortion is an output parameter that the algorithm would be able to determine given a set of parameter data, it must be calculated by alternate means in order to build the algorithm. Using the methods described in (1) and (2), a basic model that ties generator pairs and distortion together can be used to provide a baseline for datasets. This parameter information can also be obtained through board files, SPICE simulations and PCB design properties, such as the resistivity of the traces.

To perform a prediction on a circuit design, the training data from a master data file is used to construct the regression model and its coefficients. This is currently performed using the Scikit-Learn module available in Python. Data from a new PCB design is pulled to show how the data in the master training data file was and is then stored in a matrix to sequentially hold all traces as data points. This will be used as a validation data set. Each data point is then cycled through the regression model, where a prediction for the EMI distortion is made. Using this prediction, a maximization is then performed using the data point parameter values and model coefficients to sweep a generator trace length value up to the maximum length of the trace. The optimized value of the generator is then inserted with the rest of the parameter data into the matrix, which can then be output for review.

6. Algorithm Testing

Using mathematically generated EMI distortion data has a challenging issue. Because it is already related to the generator length, the results of creating a machine learning model would heavily favor the influence of values selected as the generator trace lengths. This is a display of high correlation (which due to the existence of mathematical relationships could even allow these parameters to be considered linearly dependent), which must be rectified to legitimize the model.

To address the potentially large coefficients that would dominate the computation, L1 Lasso (Least Absolute Shrinkage and Selection Operator) Regression can be implemented [5]. This type of model regularization takes the form

$$RSS_{LASSO}(w,b) = \sum_{i=1}^{N}(yi - (w \times xi + b))^2 + \alpha \sum_{j=1}^{p} |w_j| \quad (4)$$

and can help to linearly reduce extremely large model coefficients while allowing the smaller coefficients to have a greater influence on the model. This is handled by the tuning parameter α, where a value of 0 results in no coefficient reduction, while increased values contribute more to the model penalty. Care must be taken to find an appropriate value for α. Multivariate polynomial regression models will contain many coefficients, where a large amount can be acceptably reduced to zero by LASSO regression to prevent too much variance and to help limit dominating coefficients. However, α must not be too large to further overbias the model by reducing too many coefficients that have a medium impact to zero. Future testing will involve a selection process for a suitable α value based on the nature of the coefficients generated by training data.

It is necessary that when using a machine learning algorithm to optimize data points, the quality of the model is tested. This is to assure that the model can be applied as a reliable

solution to the problem it is attempting to solve. To assess the quality of the machine learning algorithm, a form of cross-validation will be performed on the training data list. The data will be divided into evenly sized partitions, where one of the partitions will be chosen to be the validation data set (the data that is input as a new PCB design to be optimized). This data set will then be run through the algorithm using the remaining partitions as the training sets. The results of the optimization will be computed as a loss quantity relative to the actual EMI distortion output data associated with the values in the validation set. If the loss is consistently low, the model is considered reliable. This process is repeated using every partition as the validation set at least once. An overall loss of the model can then be computed.

7. Experimental Results

Ultimately the most valuable data that can be used to construct the model is the data generated by actually testing the distortion effects of generator-receptor pairings in a PCB design. The results from the previously discussed generator receptor testing contain the determined distortion percentage of a protected sine wave. The parameter information from the PCB design and test results can be used as a basis for the rest of the data in the model.

Fig. 4. 4-Fold partition cross validation.

PCB design files containing the architecture for the tested traces were used to extract the generator and protected trace dimensions. This is critical to mathematically calculate their lengths. The differing design between the protected traces and the generator-receptor pairs is not fully accounted for yet, but the generators being substantially longer than the protected traces in a range of physical designs would allow length to be a contributing factor. This data would still be a valid consideration for building the model.

Other parameter information is easily obtainable from the test. The tested frequency range was 3MHz-15MHz, providing a constant range parameter value. Number of node connections and IC signal distortion is irrelevant for this design, causing this data to be primarily influenced by the generator pair and trace lengths. This will lead to a larger than

expected sparsity in the machine learning coefficients due to the data having a stronger linear relationship. This would be an opportune condition to use LASSO regression to assist with limiting the dominant coefficients.

The data was input into Python code and used to generate a multivariate polynomial regression model of degree 4. The discussed parameters were entered into the data matrix, with the lengths pulled from the design files converted to millimeters. The resulting distortion amounts from the test data were converted from percentages to a decibel representation of signal reduction. The resulting machine learning model contained 125 coefficients, 68 of which were zero, making the model with the current data 54.4 percent sparse.

To do a preliminary confirmation on the model coefficients, the first trace design values were used as a test point. The polynomial model was able to correctly provide the associated EMI distortion signal reduction value that the data was initially provided with. The data can be found in Table 1.

Table 1. Data and testing.

	Node Connections	Frequency Range	Trace Length	Generator Length	EMI Distortion
x_1	1.00e+00	1.20e+07	3.75e+01	1.28e+03	-2.45dB
x_2	1.00e+00	1.20e+07	2.55e+01	7.37e+03	-1.75dB
x_3	1.00e+00	1.20e+07	3.77e+01	2.60e+03	-1.01dB
x_4	1.00e+00	1.20e+07	4.00e+01	2.31e+03	-2.49dB
Test	1.00e+00	1.20e+07	3.75e+01	1.28e+03	-2.45dB

To continue improving the model, more data must be obtained to build the training data set. Data from previous test results is very useful to start with, however having PCB designs that can test the other parameters would be needed to incorporate the model parameters in a more meaningful way. The test does not use traces in an actual design, so no information can be gained about the number of node connections and how it relates to EMI distortion. More data can be pulled from additional PCB designs and be simulated for electrical property information, but this still relies on calculating a predicted EMI. The ideal way to obtain more training data would be to include EMI test results using a completed PCB design, requiring that each probed location have its original signal be compared to the resulting signal when the generator traces are activated. This would take more resources to obtain, but would significantly improve the quality of the data.

8. Conclusion

A method to protect PCB designs from being reverse engineered is developed, following principles of circuit obfuscation. This method will utilize the placement of transformative and nonconductive material vias incorporated into the circuit design to provide automatic and effective protection through apparent circuit possibility complexity. A particular

application of this via placement has been tested using augmented generator traces to inflict EMI into functional signals. Future testing will use machine learning to optimize the placement and characteristics of generator pairs.

References

1. K. S. McFarland *et al.*, NuTeV measurements, hep-ex/0205080. Grand, Joe. "Printed circuit board deconstruction techniques," *8th USENIX Workshop on Offensive Technologies (WOOT 14*), 2014.
2. S. Chen *et al*., "Chip-level anti-reverse engineering using transformable interconnects," *IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFTS)*, 2015.
3. S. Chen and L. Wang, "Transformable electronics implantation in ROM for anti-reverse engineering," *International Journal of High Speed Electronics and Systems*, 2019.
4. H. H. Park *et al*., "Prediction of radiated EMI from PCB excited by switching noise of IC," *Microwave and Optical Technology Letters* **51**(10), 2262–2266 (2009).
5. R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(1), 267–288 (1996).