QFlow: Quantitative Information Flow for Security-Aware Hardware Design in Verilog

Lennart M. Reimann, Luca Hanel, Dominik Sisejkovic, Farhad Merchant and Rainer Leupers
Institute for Communication Technologies and Embedded Systems, RWTH Aachen University, Germany
{lennart.reimann, hanel, sisejkovic, merchant, leupers}@ice.rwth-aachen.de

Abstract—The enormous amount of code required to design modern hardware implementations often leads to critical vulnerabilities being overlooked. Especially vulnerabilities that compromise the confidentiality of sensitive data, such as cryptographic keys, have a major impact on the trustworthiness of an entire system. Information flow analysis can elaborate whether information from sensitive signals flows towards outputs or untrusted components of the system. But most of these analytical strategies rely on the non-interference property, stating that the untrusted targets must not be influenced by the source’s data, which is shown to be too inflexible for many applications. To address this issue, there are approaches to quantify the information flow between components such that insignificant leakage can be neglected. Due to the high computational complexity of this quantification, approximations are needed, which introduce mispredictions. To tackle those limitations, we reformulate the approximations. Further, we propose a tool QFlow with a higher detection rate than previous tools. It can be used by non-experienced users to identify data leakages in hardware designs, thus facilitating a security-aware design process.

Index Terms—hardware security, hardware Trojans, quantitative information flow, vulnerability, confidentiality

I. INTRODUCTION

Many security issues are either caused accidentally by inexperienced hardware designers or by malicious modifications, such as hardware Trojans [1] [2]. In the rest of this work, those issues are referred to as vulnerabilities. The vulnerabilities should be identified and removed at an early design stage, as later modifications result in higher costs or a longer time-to-market. They can be identified using suitable test models, but the required test models become more computationally extensive with the increasing design complexity [3], thus not all test cases can be elaborated in a reasonable time. Additionally, theorem provers [4], property checkers, and formal approaches give more certainty, but often require special expertise to be used properly and suffer from scalability issues.

One of the critical features of hardware security is the confidentiality of data. Vulnerabilities endangering the confidentiality of sensitive signals, such as cryptographic keys, can be implemented easily and stay undetected in common test cases, as the trigger might not be present in the program code and data. Therefore, we focus on protecting the confidentiality in this work. Information flow analysis is an evolving research area and used as a method to detect hardware vulnerabilities concerning the confidentiality of data carried by digital circuits. A variety of solutions exist working on virtual prototypes [5] or Register-Transfer Level (RTL) [6]–[8] to analyze the flow of information of marked signals, either dynamically (tracking) or statically (analysis). As dynamic elaborations analyze the information flow at runtime, they can only guarantee the security of a hardware design for given test cases [5]. Most of these tools rely on analyzing the flow of information for the non-interference property. But this property, which forbids the communication between trusted and untrusted hardware components, can often not be implemented, as many applications rely on such a communication. State-of-the-art methods, such as QIF-Verilog [9], use Quantitative Information Flow (QIF) analysis, which allows a quantification of the actual leakage using developed metrics that have been shown to be suitable to detect leakages in digital systems [10] [11]. This analysis allows a new classification of leakage paths supporting the identification of data leakages. However, in this work, QIF-Verilog is shown to be unreliable in the identification of certain design vulnerabilities. Therefore, we introduce a more suitable methodology, QFlow, capable of detecting even the new vulnerabilities.

The major contributions of this paper are: i) An operational tool that uses a suitable QIF metric, called QModel, to quantify the trustworthiness of a Verilog hardware description regarding the protection of its sensitive signals, e.g., cryptographic keys or user data, ii) A secure approximation of state-of-the-art QIF formulas that removes the possibility of false negative predictions, while only enabling false positives, as shown empirically and iii) User-adjustable input probabilities for the information theory equations on bit level.

A. Attack Model

We focus on vulnerabilities that are implemented during the RTL-design process. Those vulnerabilities can be exploited by an adversary after production. In this work, we assume that the attacker can observe the outputs and the non-secret inputs of the selected hardware modules at a random moment in time. These outputs might leak signals carrying sensitive information, such as user data, via leakage paths. Whether those observations are obtained by physical access to the design or via other untrusted hardware in the System-on-Chip (SoC) is not considered. Additionally, the intruder cannot set any of the input values of the module under attack. During the attack, the complete design structure is known to the adversary.

B. Information Flow Analysis

The designer is interested in protecting certain parts of the hardware carrying sensitive data. For this purpose, previous works used labels to classify the sensitivity of hardware models and the data that is carried in them [12]. In most information flow models the system of interest is separated into two partitions: High (H) and Low (L). The H label is commonly used to describe trusted hardware components processing sensitive data. When labeling the hardware components, the trusted components would be labeled H,
the remaining components are labeled L. These are abstract labels, but the labels are commonly applied by marking the component’s hardware description with their respective label [6]. The labels H and L are propagated throughout the system related to the dependency of the signals.

C. Quantitative Information Flow

In many previous works [13] [14], the vulnerability, a metric for the weakness or simplicity to guess the secret, has shown to be a reliable function when quantifying the information flow. The Bayes Vulnerability represents a special case of the g-vulnerability, when the adversary has only a single guess of the secret after one observation of the outputs and is only rewarded for a correct guess [15] [10]. For a secret H with a probability distribution of \( \pi_H \), the vulnerability is

\[
V_1[\pi_H] = \max_{h \in H} \pi_h,
\]

(1)

where \( \pi_h \) is the probability of the symbol \( h \in H \). In this work, we are interested in the leakage caused by an information flow from a secret H to the outputs of a system O. Therefore, the Posterior Bayes Vulnerability \( V_1[\pi_H \triangleright C] \) (PBV) has to be considered as well. The posterior vulnerability shows the vulnerability after observing the outputs O, if the secret bits H have been applied to the deterministic channel C (\( \pi_H \triangleright C \)):

\[
V_1[\pi_H \triangleright C] = \sum_{o \in O} \max_{h \in H} J_{o,h}.
\]

(2)

Here, \( J \) represents the joint distribution for the variables in the index. The channel is an abstract definition of the hardware system. When combining the two Bayes Vulnerabilities, the so-called Multiplicative Bayes Leakage \( L_1^\times \) is computed as:

\[
L_1^\times := \frac{V_1[\pi_H \triangleright C]}{V_1[\pi]}.
\]

(3)

As the leakage computation for a complete hardware description would be infeasible, approximations are needed.

III. QFlow

A. Basic Functionality

Our QIF-metrics are used to quantify the information flow from the 'High' marked sources, via all the operations that are applied on the secret, to the 'Low' targets. All output ports of the top design are automatically marked as 'Low'. In contrast to QIF-Verilog, we preprocess the abstract syntax tree to allow a bit-wise analysis to track the information flow in more detail. We compute the leakages in several steps. First we build channels consisting of Boolean equations representing the partitioned hardware. Next, the input probabilities and PBV for the channels are computed and combined with the input leakages of the channel to compute the total leakages for every single secret bit.

B. Approximation Assumptions

Certain assumptions are needed to reduce the amount of computations needed to determine leakage values for the secrets: i) The secret and known bits are independent of each other, ii) Compute Probabilities: The inputs of every channel are independent of each other, iii) Compute Posterior Vulnerability: The inputs of every channel are independent of each other. Additionally for COMPARISON, ADDITION, and SUBTRACTION operations, the high inputs are assumed to be uniformly distributed and iv) Compute Leakage: The dependency of input leakages is appraised with Eq. (3). As the low inputs can be observed by the adversary as well, they need to be integrated into the equation for the PBV.

C. QModel

The assumptions and appraisals lead to our mathematical model, called QModel, implemented in our tool QFlow. This model is further explained below. As mentioned before, we modified the equation for the PBV (Eq. (2)) to include the low input signals L. They are added to the equation as additional observable parameters—similar to the outputs O:

\[
V_1[\pi \triangleright C | L] := V_1[H | L, O] := \sum_{o \in O} \max_{h \in H} J_{o,l,h},
\]

(4)

with the joint probability distribution \( J_{o,l,h} \).

\[
J_{o,h,l} = \pi_{o|h,l} \cdot \pi_{h,l}.
\]

(5)

A feasible way to compute the leakage in a channel cascade is needed. Thus, we derived the following statement for the leakage \( L_{X \rightarrow Z} \) (in bit) of a Markov chain:

\[
L_{X \rightarrow Z} := L_{X \rightarrow Y} \cdot \frac{L_{Y \rightarrow Z}}{L_{Y \rightarrow Z,\text{max}}}.
\]

(6)

For a Markov chain of channels \( X \rightarrow Y \rightarrow Z \), the input values \( X \) are converted to the output \( Z \) with \( Y \) as an intermediate value. The data processing inequality states that no processing of \( Y \) can increase the information that \( Y \) has about \( X \) [15]. Thus, we weigh the leakage of channel \( X \rightarrow Y \) with the leakage of the second channel, scaled with its maximum possible leakage. When inserting the multiplicative Bayes Leakage (Eq. (3)) and the maximum leakage for the second channel, the reciprocal of the Prior Bayes Vulnerability (Eq. (1)), we derive the following equation:

\[
L_{X \rightarrow Z} = L_{X \rightarrow Y} \cdot \frac{V_1[\pi_Y \triangleright C_{Y \rightarrow Z}]}{V_1[\pi_Y]}
= L_{X \rightarrow Y} \cdot V_1[\pi_Y \triangleright C_{Y \rightarrow Z}].
\]

(7)

The input leakage into a channel (now in bit) can be further approximated as stated before in the Section III-B using [16]

\[
L \leq \sum_{i \in \text{Inputs}} L_i,
\]

(8)

resulting in our final equation for the leakage of a secret bit \( H_j \) in a channel cascade \( L_{C,H_j} \). The probability distribution of the secret channel inputs is represented by \( \pi_{H1} \) and is used to compute the overall PBV of that channel:

\[
L_{C,H_j} := \sum_{i \in \text{Inputs}} L_{i,H_j} \cdot V_1(\pi_{H1} \triangleright C).
\]

(9)

D. Toolflow

The general toolflow of QFlow is illustrated in Fig. 1:

1) ParseArgs: Some parameters and the Verilog code are passed to the program. The signal that is supposed to stay secret is marked as 'High'.

2) QIF-Parser: The QIF-parser integrates the open-source tool PyVerilog, which is responsible for parsing Verilog and returning a graph of binds. A bind in such a graph mostly represents a single line of RTL code with terminal symbols,
such as signals and constants as its leaves. As we intend to work on the bit level of the design, we need to process the bind tree. Then, each tree has an output bit as the root and only input signals and constants can be leaves. An example for such a tree can be found in Fig. 2a. For the given circuit in Fig. 2a the tree structure is illustrated for the first bit of the output in Fig. 2b. The leaves (green) consist of the inputs of the circuit and constants. Operations (yellow) are connected to their outputs (orange) until an output, the root (red), is reached.

3) Compute Dependencies & Merge: Afterwards, the tool computes the dependencies and starts merging nodes. The dependencies are needed to find loops in the tree structure to avoid merging them infinitely. Merging is done to increase the channel sizes, thus reducing the number of channels that are cascaded. Every cascade of channels introduces an error in the computed leakage, due to the approximations done with the equations presented before. A negative error can be introduced due to the assumption of independent channel inputs, whereby the simple addition of input leakages introduces a positive error. In this paper, it is shown empirically that the positive error outweighs the negative one. The merging is stopped when the maximum number of input bits in the Boolean expression (max_channel_inputs) is reached. This is done to reduce the complexity for the next step, computing the probabilities. Furthermore, sequential branches are not merged as this may result in a security risk and misinterpretation of hardware vulnerabilities, which introduces an additional positive error. The example’s tree was merged with max_channel_inputs=3, as shown in Fig. 2c.

4) Compute Probabilities: The merged trees are forwarded to the ‘Compute Probabilities’ function. Here, the output probabilities of every channel input bit are computed. The probability of the outputs is calculated by multiplying the input probabilities for a given case because of the assumed independence. An example for the computation of the probabilities is shown in Fig. 2c with uniform input probabilities.

5) Compute Vulnerability & Leakage: All the computed probabilities are written into the respective nodes. Next, it is possible to compute the posterior vulnerabilities for every channel using Eq. 4. All the example’s computed probabilities are illustrated in Fig. 2c. For the computation of the PBV, the joint probability distributions need to be computed with Eq. 5 for every channel. Furthermore, the leakage of the input bits is set for every secret bit. If multiple inputs hold information about a secret bit, their leakage is added beforehand to appraise possible dependencies. From that point on, the leakage of every secret bit is multiplied with the PBV of the current channel. For every output bit, the leakages for all secret bits that influence it are computed by using the Eq. 9 over all channels starting at the leaves. Finally, the total leakage for every secret bit is computed by adding their leakage from every output bit.

6) Compare with two Thresholds: Now that the leakage values for the different secret input bits are computed, we need to determine, which of the data paths are actually a vulnerability in the system. We determine a threshold and compare our computed value for every secret bit to it.

7) Classification: After the accumulated leakages for every bit that is labeled with ‘High’ is computed and they have been compared to the thresholds, the leakage paths can be written out. This is done in three groups: Paths that leak (higher than detection threshold), paths that might leak (warning threshold), and paths that do not leak (below both thresholds). For the example, this is shown with the red connections in Fig. 2d.

IV. EVALUATION

For the evaluation, multiple open-source benchmarks were used to show the efficiency and flexibility of QFlow in detecting security vulnerabilities concerning the confidentiality of data. During the design process, the hardware designer is capable of elaborating the security of certain signals by marking them with ‘High’ inside his code. A second parameter that we alternate during the experimentation is the probability of the high input bits. This is done to show the influence of the probability on the leakage and illustrates the capability of the tool to emphasize the detection of hardware Trojans using controllable triggers. The same benchmarks are used to elaborate QIF-Verilog as well, thus they allow a comparison. In the beginning, the required thresholds were determined by

![Fig. 1: General toolflow of QFlow.](image1)

![Fig. 2: Example for the computational steps for a small circuit.](image2)
applying QFlow to two pipelined rounds of AES. This was done for uniformly distributed inputs, as they allow the highest leakage, leading to a more general threshold, while varying max_channel_inputs. For most of the elaborations, AES and RSA benchmarks from Trust-Hub [17] are used, which offer cryptographic accelerators in Verilog (and VHDL [18]). They offer a variety of Trojans using either triggers or continuously write out the keys over additional output ports. Furthermore, we implemented additional benchmarks to prove QFlow’s functionality compared to QIF-Verilog.

A. Results

At first, we analyzed a single AES benchmark for varying high input probabilities to illustrate the flexibility of QFlow and show the meaning of leakage. For the analysis, the probability of the 128-bit key was altered from 0 to 1 for all secret bits equally. The results (Fig. 3a) show that although the leakage for the system is 0 for the higher and lower probabilities, the attacker can still guess the keys as the Prior Bayes Vulnerability is at 1 bit. Even a solid hardware implementation cannot protect an insecure secret.

As mentioned before, the threshold was determined by applying QFlow to a Verilog implementation of two pipelined rounds of AES (by reducing one of the AES benchmarks). Two complete rounds of AES are supposed to lead to a full diffusion [19]. Fig. 3 shows the minimum and average value of the 128-bit leakages for the keys with a varying max_channel_inputs value. As expected, the leakage drops, when more operations are merged into a single channel, reducing the error introduced by Eq. 5 and the assumed independence of inputs. The smallest mean and min leakages that were computed were chosen as the threshold for the warning (min, 2.89154⋅10^{-3}) and threat (mean, 1.53939⋅10^{-3}) for the toolflow. The leakage did not drop after setting the maximum number of input values during the merge to 5. Thus, the value for max_channel_inputs is set to 5 for all following elaborations. After the threshold was determined, the first analyses on the AES benchmarks were conducted. The results of these experiments can be seen in Fig. 3b illustrating the computed leakage for two different Trojans over the 128-bit key. The T-1200 benchmark leaks the first 8 bit of the key, which was XOR-ed with a value that can be determined by the intruder value. Thus, the XOR is not reducing the information content. Additionally, the Trojan is triggered continuously. applying QFlow to two pipelined rounds of AES. This was done for uniformly distributed inputs, as they allow the highest leakage, leading to a more general threshold, while varying max_channel_inputs. For most of the elaborations, AES and RSA benchmarks from Trust-Hub [17] are used, which offer cryptographic accelerators in Verilog (and VHDL [18]). They offer a variety of Trojans using either triggers or continuously write out the keys over additional output ports. Furthermore, we implemented additional benchmarks to prove QFlow’s functionality compared to QIF-Verilog.

A. Results

At first, we analyzed a single AES benchmark for varying high input probabilities to illustrate the flexibility of QFlow and show the meaning of leakage. For the analysis, the probability of the 128-bit key was altered from 0 to 1 for all secret bits equally. The results (Fig. 3a) show that although the leakage for the system is 0 for the higher and lower probabilities, the attacker can still guess the keys as the Prior Bayes Vulnerability is at 1 bit. Even a solid hardware implementation cannot protect an insecure secret.

As mentioned before, the threshold was determined by applying QFlow to a Verilog implementation of two pipelined rounds of AES (by reducing one of the AES benchmarks). Two complete rounds of AES are supposed to lead to a full diffusion [19]. Fig. 3 shows the minimum and average value of the 128-bit leakages for the keys with a varying max_channel_inputs value. As expected, the leakage drops, when more operations are merged into a single channel, reducing the error introduced by Eq. 5 and the assumed independence of inputs. The smallest mean and min leakages that were computed were chosen as the threshold for the warning (min, 2.89154⋅10^{-3}) and threat (mean, 1.53939⋅10^{-3}) for the toolflow. The leakage did not drop after setting the maximum number of input values during the merge to 5. Thus, the value for max_channel_inputs is set to 5 for all following elaborations. After the threshold was determined, the first analyses on the AES benchmarks were conducted. The results of these experiments can be seen in Fig. 3b illustrating the computed leakage for two different Trojans over the 128-bit key. The T-1200 benchmark leaks the first 8 bit of the key, which was XOR-ed with a value that can be determined by the intruder value. Thus, the XOR is not reducing the information content. Additionally, the Trojan is triggered continuously.
triggers that make the circuit most vulnerable. This reduces the computed leakage compared to the actual leakage for highly unlikely leakage paths. Furthermore, hardware Trojans or unintentional vulnerabilities that leak a low amount of information over a longer period of time might not be detected.

V. CONCLUSION

In this publication, we introduced QFlow, a tool allowing the hardware designer to create a more security-aware design process using Quantitative Information Flow. It was shown that by using a suitable approximation, the hardware design can be separated into several channels to reduce the computational complexity, when independence of the channel’s inputs is assumed. However, this dependency needs to be considered for merging leakage paths. The tool was proven to be more reliable for the defined attack model than the state-of-the-art tools. A simple usage and the acceptable computation time, support the claim that QFlow allows a security-aware design process that can be conducted by inexperienced users.

REFERENCES

[1] S. Bhunia et al., “The hardware trojan war: Attacks, myths, and defenses,” Springer International Publishing, 2018.
[2] D. Šišeković et al., “Scaling logic locking schemes to multi-module hardware designs,” in Architecture of Computing Systems – ARCS 2020.
[3] T. Marchok et al., “Complexity of sequential ATPG,” 1995, pp. 252–261.
[4] X. Guo et al., “Scalable SoC trust verification using integrated theorem proving and model checking,” in 2016 IEEE HOST, 2016, pp. 124–129.
[5] M. Hassan et al., “Early SoC security validation by VP-based static information flow analysis,” in IEEE/ACM ICCAD, 2017, pp. 400–407.
[6] D. Zhang et al., “A hardware design language for timing-sensitive information-flow security,” SIGARCH Comput. Archit. News, 2015.
[7] X. Li et al., “Sapper: A language for hardware-level security policy enforcement,” SIGPLAN Not., 2014.
[8] X. Li et al., “Caisson: A hardware description language for secure information flow,” SIGPLAN Not., 2011.
[9] X. Guo et al., “QIF-Verilog: Quantitative information-flow based hardware description languages for pre-silicon security assessment,” HOST ’19.
[10] G. Smith, “On the foundations of quantitative information flow,” in Foundations of Software Science and Computational Structures, 2009.
[11] M. S. Alvim et al., “Measuring Information Leakage using Generalized Gain Functions,” in Computer Security Foundations. IEEE, 2012.
[12] A. Ferraiuolo et al., “Verification of a practical hardware security architecture through static information flow analysis,” SIGARCH Comput. Archit. News, 2017.
[13] S. H. Hussein, “Refining a quantitative information flow metric,” 2012 5th NTMS, May 2012.
[14] J. Zhu et al., “Poster: on quantitative information flow metrics,” 2011.
[15] M. S. Alvim et al., The Science of Quantitative Information Flow, ser. Information Security and Cryptography, 2020.
[16] M. Boreale, “Quantifying information leakage in process calculi,” in Automata, Languages and Programming, M. Bugliesi et al., Eds., 2006.
[17] B. Shaky et al., “Benchmarking of hardware Trojans and maliciously affected circuits,” Journal of Hardware and Systems Security, 2017.
[18] “Vhd2Vl,” visited 2021-05-27. [Online]. Available: https://github.com/ldolinski/vhd2vl
[19] J. Daemen et al., The Design of Rijndael: AES - The Advanced Encryption Standard, 1st ed. Springer, 2002.
[20] H. Salmani et al., “On design vulnerability analysis and trust benchmarks development,” in 2013 IEEE 31st ICCD, 2013.

| TABLE I: AES- and RSA-Trojan benchmark leakages |  |
|-----------------------------------------------|---|
| **Benchmark** | **Detected #Actual Leakage** | **Avg. Det. Leakage** | **#Detected #Actual** | **#Detected #Actual** | **#Detected #Actual** | **#Detected #Actual** | **Time (s)** |
| AES-T100 | 8/8 | 1 | 0/0 | 0/- | 120/120 | 2.66 · 10⁻⁴ | 246 |
| AES-T200 | 8/8 | 1 | 0/0 | 0/- | 120/120 | 2.66 · 10⁻⁴ | 214 |
| AES-T400 | 128/128 | 0.183 | 0/0 | 0/0 | 0/0 | 0/0 | - |
| AES-T700 | 8/8 | 1 | 0/0 | 0/- | 120/120 | 2.66 · 10⁻⁴ | 236 |
| AES-T800 | 8/8 | 1 | 0/0 | 0/- | 120/120 | 2.66 · 10⁻⁴ | 232 |
| AES-T900 | 8/8 | 1 | 0/0 | 0/- | 120/120 | 2.66 · 10⁻⁴ | 233 |
| AES-T1000 | 8/8 | 1 | 0/0 | 0/- | 120/120 | 2.66 · 10⁻⁴ | 232 |
| AES-T1100 | 8/8 | 1 | 0/0 | 0/- | 120/120 | 2.66 · 10⁻⁴ | 238 |
| AES-T1200 | 8/8 | 1 | 0/0 | 0/- | 120/120 | 2.66 · 10⁻⁴ | 233 |
| AES-T1600 | 128/128 | 0.222 | 0/0 | 0/0 | 0/0 | 0/0 | - |
| AES-T1700 | 128/128 | 0.295 | 0/0 | 0/0 | 0/0 | 0/0 | - |
| RSA-T100 | 33/32 | 0.5 | 1/0.023 | 1/0.006 | 30/32 | 6.3 · 10⁻² | 196 |
| RSA-T300 | 33/32 | 0.5 | 0/0.023 | 1/0.023 | 30/32 | 6.3 · 10⁻² | 191 |

![Fig. 5: New benchmarks to show the vulnerabilities in QIF-Verilog. Three Trojans leaking 64 bit.](image)