

CleverHans: Precise Fine-grained Side-channel Information Leakage Quantification in Binaries

Anonymous, v10

Abstract—Side-channel attacks allow attackers to infer some sensitive information based on non-functional characteristics. Existing works on address-based side-channel detection can provide a list of potential side-channel leakage sites. However, they 1) only consider the “average” information leakage which often neglect severe attack scenarios, 2) cannot accurately report the sum of multiple information leakages from various sites at one time.

To overcome the above limitations, we propose a novel method to quantify the information leakage more precisely. The key insight is **FIXME:** XXXXX. We have developed a tool called CleverHans, which can not only find the side-channel vulnerabilities but also estimate how many bits are actually leaked. CleverHans works in three steps. First, the application is executed to record the trace. Second, CleverHans runs the instruction level symbolic execution on the top of the execution trace. CleverHans will find side-channel information leakages and model each leakage as one unique math constraint. Finally, CleverHans will classify those constraints into independent multiple groups and run the Markov Chain Monte Carlo to estimate the information leakage. CleverHans can report a very fine-grained vulnerability result compared to the existing tools.

We applied the tool on OpenSSL, MbedTLS and libjpeg and found several serious side-channel vulnerabilities. Compared with previous research, the results are surprisingly different that most of the reported vulnerabilities are hard to exploit in practice.

I. INTRODUCTION

Side channels are hardly avoided in modern computer systems as the sensitive information may be leaked by kinds of neglected behaviors, such as power, electromagnetic radiation and even sound []. Among them, software-based side-channel vulnerabilities are especially common and have been studied for years []. The vulnerability results from vulnerable software codes and shared hardware components. By observing the outputs or hardware behaviors, attackers could infer the program execution flow with secrets and guess the secrets such as encryption keys [].

Various countermeasures have been proposed to defend against software-based side-channel attacks by reducing the shared resources and eliminating the leak paths [1]–[3]. In addition to runtime overheads, they are limited to specific hardwares. So it is imperative to discover and fix side channel vulnerabilities in advance and automatic detection methods are proposed recently [].

In general, both static analysis and dynamic analysis can be used to detect software-based side-channel vulnerabilities. Static methods include **FIXME:** .xxx and dynamic methods include **FIXME:** .xxx. They help to find a list of hidden vulnerabilities in real-world software, but fail to report how severe a potential leakage site could be. Many of the reported

vulnerabilities are typically hard to be exploited and leak very little information, e.g., one bit of the length of the key [2]. For example, the vulnerabilities in OpenSSL library are assigned with low severity and not considered in the threat model now [?]. It is worthwhile to have a tool to tell how much information is actually leaked and help to fix them.

To assess side-channel vulnerabilities, we need a proper quantification metric. Tools based on static analysis with abstract interpretation can give an upper bound, which is useful to justify whether the implementation is secure enough. However, they can't indicate the leakage is severe or not due to the over-approximation []. The “average” information leakage assumes that the target program will have **variable** sensitive information when the attacker launches the attack. But for real-world attacks, an adversary may run the target problem again and over again with **fixed** unknown sensitive information such as the key. The dynamic methods take another way with a concrete input and run the program actually. Although they are very precise, they suffer from the incompleteness of testing and trade-off of performance and accuracy []. **FIXME:** *why DATA example is commented out below.*

Additionally, open source libraries may have multiple leakage sites, which can be exploited for attackers at one time [4]–[6]. It is necessary to know how much information is actually leaked in total, but no existing tools can practically estimate multiple leakage sites in real-world libraries. Simply adding the results together gets a very rough upper bound of the total information leakage if those leakages aren't independent.

To overcome the above limitations, we propose a novel method to quantify the information leakage more precisely. In general, we quantify the leaked information as the uncertain input space based on attackers' observation. Before the attack, the adversary has a big but finite input space. Every time the adversary observes one leakage site, he can eliminate some potential input and reduce the size of the input space. In extreme cases, if the size of the input space reduces to one, the adversary can determine the input information uniquely, which means all information (e.g., the whole secret key) is leaked. Further, we turn the quantification into a model counting problem [?] and solve it by Markov Chain Monte Carlo approximate counting [?]. In this way, we quantify the information leakage more precisely.

More specifically, we build a tool called CleverHans,¹ which could discover and estimate those potential information

¹Clever Hans is a horse that can “count”. Our tool uses an advanced method to count the number of leaked bits through side channel.

leakage sites as well as how many bits they can leak. We assume that adversaries can exploit different control-flow transfers and data-access patterns when the program processes different sensitive data. First, we collect the dynamic execution trace for each input of the target libraries and then run symbolic execution on the traces. In this way, we model each side-channel leakage as a math formula. The sensitive input is divided into several independent bytes and each byte is regarded as a unique symbol. Those formulas can precisely model every the side-channel vulnerability. Then we extend the problem to multiple leakages and related leakages and introduce a monte carlo sampling method to estimate the total information leakage. In fact, if the application has a different sensitive input but still satisfies the formula, the code can still leak the same information.

We apply CleverHans on several crypto and non-crypto libraries including OpenSSL, MbedTLS and libjpeg. The experiment result confirms that CleverHans can precisely identify the pre-known vulnerabilities, reports how much information is leaked and which byte in the original sensitive buffer is leaked. Although some of the analyzed crypto libraries have a number of side-channels, they actually leak very little information. Also, we also use the tool to analysis the two reported side-channel attack in the libjpeg library. Finally, we present new vulnerabilities. With the help of CleverHans, we confirm those vulnerabilities are easily to be exploited. Our results are superisingly different compared to previous results and much more useful in practice.

In summary, we make the following contributions:

- We propose a novel method that can quantify fine-grained side-channel information leakages. We model each information leakage vulnerability as math formulas and mutiple side-channel vulnerabilities can be seen as the disjunctions of those formulas, which precisely models the program semantics.
- We transfer the information quantification problem into a probabily distribution problem and use the Monte Carlo sampling method to estimate the information leakage. Some initial results indicate the the sampling method suffers from the curse of dimensionality problem. We therefore design a guided sampling method and provide the corresponding error esitimate.
- We implement the proposed method into a practical tool and apply it on several real-world software. CleverHans successfully identify the address-based side-channel vulnerabilities and provide the corresponding information leakage. The information leakage result provide the detailed information that help developers to fix the vulnerability.

II. BACKGROUND

In this section, we first present a basic introduction to the software-based side-channel attacks. Those attacks are what we attempt to study in the paper. After that, we will show existing methods on side-channel detection and quantification.

A. Software-based Side-Channels

Software-based side-channels leaks sensitive information unconsciously through different behaviors. Fundamentally, those differences were caused by the memory hierarchy design in modern computer systems. When the CPU fetches the data, it first searches the cache, which stores copies of the data that was frequently used in main memory. If the data doesn't exist in the cache, the CPU will read the data from the main memory (RAM). According to the layer that causes the side-channels, we classify the software-based side channels into two categories: cache-based side-channel attacks and memory-based side-channel attacks.

1) *Cache-based Attacks*: In general, the cached-based side-channels seek information relying on the time differences between the cache miss and the cache hit. Here we introduce two frequently used cache attacks: PRIME+PROBE and FLUSH+RELOAD.

PRIME+PROBE targets a single cache set. It has two phases. During the “prime” phase, the attacker fills the cache set with his data. In the second “probe” phase, the attacker re-accesses the cache set again. If the victim accesses the cache set and evicts part of the data, the attacker will experience a slow measurement. If not, it will be fast. By knowing which cache set the target program that accesses, the attacker can infer parts of the sensitive information.

FLUSH+RELOAD targets a single cache line. It requires the attacker and victim share the same memory. It also has two phases. During the “flush” stage, the attacker flushes the “monitored memory” from the cache. Then the attacker waits for the victim to access the memory. In the next phase, the attacker reload the “monitored memory”. If the time is short, which indicates there is a cache hit and the victim reloads the memory before. On the other hand, the time will be longer since the CPU need to reload the memory into the cache line.

2) *Memory-based Attack*: Memory-based side-channel attack [5] exploits the different behaviors when the program accesses different page tables. For example, the controlled-channel attack [5], which works in the kernel space, can infer the sensitive data in the shielding systems by observing the page fault sequences after restricting some code and data pages.

After examining the memory-based side-channels attack, we believe the root cause of those attacks are secret-dependent memory access and control flow transfers. **FIXME:** *This paragraph is confusing. Secret-dependent memory access and control flow transfers can be cache-based side channels. We need further develop this so it is not a surprise when it is mentioned in the next subsection.*

B. Information Leakage Quantification

FIXME: *Section (II.B), this subsection: Clarify mutual information, min-entropy, and maximal leakage. How many types?*

Given an event e which occurs with the probability $P(e)$, if the event e happens, then we receive

$$I = -\log_2 P(e)$$

bits of information by knowing the event e . Consider a char variable a in a C program, which has the size of one byte (8 bits). It ranges from 0 – 255. Assume a has the uniform distribution. If at one time we observe that a equals 1, the probability of this observation is $\frac{1}{256}$. So the information we get is $-\log(\frac{1}{256}) = 8$ bits, which is exactly the size of the char variable in the C program.

Some existing work on information leakage quantification is based on mutual information or min-entropy [7]. In these frameworks, the input sensitive information K is viewed as a random variable. Let k_i be one of the possible value of K . The Shannon entropy $H(K)$ is defined as

$$H(K) = - \sum_{k_i \in K} P(k_i) \log_2 P(k_i)$$

The Shannon entropy can be used to quantify the initial uncertainty about the sensitive information. Suppose a program with some sensitive input K , an adversary has some observations (O) through some side channel. In this work, the observations are referred as the secret-dependent control-flows and secret-dependent data-access patterns. **FIXME:** See the comment above, we need to explain these patterns before this point. The conditional entropy $H(K|O)$ is

$$H(K|O) = - \sum_{o_j \in O} P(o_j) \sum_{k_i \in K} P(k_i|o_j) \log_2 P(k_i|o_j)$$

Intuitively, the conditional information marks the uncertainty about K after the adversary has gained some observations (O).

Some previous work uses the mutual information $I(K; O)$ to quantify the leakage which is defined as follows:

$$Leakage = I(K; O) = \sum_{k_i \in K} \sum_{o_j \in O} P(k_i, o_j) \log_2 \frac{P(k_i, o_j)}{P(k_i)P(o_j)}$$

where $P(k_i, o_i)$ is the joint discrete distribution of K and O . Alternatively, the mutual information can also be computed with the following equation:

$$Leakage = I(K; O) = H(K) - H(K|O) = H(O) - H(O|K)$$

For a deterministic program, once the input K is fixed, the program will have the same control-flow transfers and data-access patterns. As a result, $P(k_i, o_j)$ will always equals to 1 or 0. So the conditional entropy $H(O|K)$ will equal to zero. So the leakage defined by the mutual information can be simplified into:

$$Leakage = I(K; O) = H(O)$$

In other words, once we know the distribution of those memory-access patterns. We can calculate how much information is actually leaked.

Another common method is based on the maximal leakage [1], [7], [8]. The formal information theory can prove that the maximal leakage is the upper bound of the mutual information (Channel Capacity).

$$Leakage = \log(C(O))$$

Here $C(O)$ represents the number of different observations that an attacker can have.

III. RUNNING EXAMPLE

Now we provide a concrete example to show how the two types of quantification definition works and show that how our method is different.

Maximal Leakage. Depending on the value of key, the code can run four different branches which corresponding to four different observations. Therefore, by the maximal leakage definition, the leakage equals to $\log 4 = 2$ bits.

```
1 unsigned char key = input();
2 // key = [0 ... 255]
3 if (key == 128)
4     A(); // branch 1
5 else if (key < 64)
6     B(); // branch 2
7 else if (key < 128)
8     C(); // branch 3
9 else
10    D(); // branch 4
```

Listing 1: A simple program

Mutual Information. If the key satisfies the uniform distribution, the probability of the code runs each branch can be computed with the following result.

Branch	A	B	C	D
Possibility	1/256	64/256	64/256	127/256

TABLE I: The distribution of observations

Therefore, the leakage equals to $\frac{1}{256} \log \frac{1}{256} + \frac{1}{4} \log \frac{1}{4} * 2 + \frac{127}{256} \log \frac{127}{256} = 1.7$ bits.

CleverHans. **FIXME:** What is out result here?

IV. THREAT MODEL

We consider an attacker who shares the same hardware resource with the victim. The attacker attempts to retrieve sensitive information via memory-based side-channel attacks. The attacker has no direct access to the memory or cache, but can probe the memory or cache at each program point. In reality, the attacker will face many possible obstacles, including the noisy observations, limited observations on memory or cache. But for this project, we assume the attacker can have noise-free observations. The threat model captures most of the cache-based and memory-based side channel attacks. We only consider the deterministic program for the project.

V. CHALLENGES

In this section, we articulate several challenges and existing problems in quantifying the side-channel vulnerability leakages. We describe the challenges and then briefly present the corresponding solution.

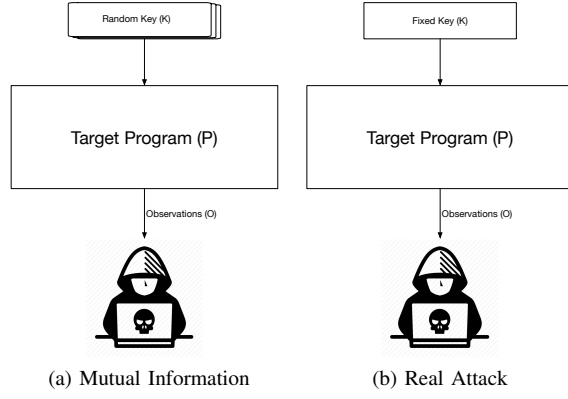


Fig. 1: some caption here. **FIXME:** redraw the figure. i will give you the instruction.

A. Challenge I: Information Leakage Definition

Existing static-based side-channel quantification works [1] define information leakage as the mutual information or the maximal leakage. These definitions provide a strong security guarantee when trying to prove a program is secure enough if their methods calculate the program leaks zero bits of information.

However, the above definition is less useful to justify the sensitive level of leakage sites. Considering the example code 1 in the previous section, if an attacker knows the code executes branch A by some observations, the attacker can know the key actually equals to 128. Suppose it is a dummy password checker, in which case the attacker can fully retrieve the password. Therefore, the total information leakage should be 8 bits, which equals to the size of unsigned char. According to the mutual definition, however, the leakage will be 1.7 bits. The maximal information leakage is 2 bits. Both approaches fail to tell how much information is actually leaked during the execution precisely.

The problem with the existing methods is that they are static-based and the input values are neglected by the previous definition. They assume the attacker runs the program multiple times with many different sensitive information as the input. Both the mutual information and the max-leakage give an “average” estimate of the information leakage. But it isn’t the typical scenario for an adversary to launch the side-channel attack. When a side-channel attack happens, the adversary wants to retrieve the sensitive information, in which case the sensitive information is fixed (e.g. AES keys). The adversary will run the attack over and over again and guess the value bit by bit. Like the previous example, the existing static method doesn’t work well in those situations.

Our Solution to Challenge I: In the project, we hope to give a very precise definition of information leakages. Suppose an attacker runs the target program multiple times with one fixed input, we want to know how much information he can infer by observing the memory access patterns. We come to

the simple slogan [7] that

$$\text{Information leakage} = \text{Initial uncertainty} - \text{Remaining uncertainty}.$$

If an adversary has zero knowledge about the input before the attack. The initial uncertainty equals to the size of the input. As for the remaining uncertainty, we come to the original definition of the information content.

We quantify the information leakage with the following definition.

Definition 1. Given a program P with the input set K , an adversary has the observation o when the input $k \in K$. We denote it as

$$P(k) = o.$$

The leakage $L_{P(k)=o}$ based on the observation is

$$L_{P(k)=o} = \log_2 |K| - \log_2 |K^o|$$

where

$$K^o = \{k' \mid k' \in K \text{ and } P(k') = o\}$$

With the new definition, if the attacker observes that the code 1 runs the branch 1, then the $K^{o^1} = \{128\}$. Therefore, the information leakage $L_{P(k)=o^1} = \log_2 256 - \log_2 1 = 8$ bits, which means the key is totally leaked. If the attacker observes the code runs other branches, the leaked information is shown in the following table.

Branch	1	2	3	4
K^o	1	64	64	127
$L_{P(k)=o}(\text{bits})$	8	2	2	1

TABLE II: The leaked information by the definition 1

With the same definition 1, if the attacker observes that the code runs branch 2, the information leakage will be 2 bits. The conclusion is consistent with the intuition. Because if the branch 2 was executed, we can know the key is less than 64. So we know the most and the second significant digits of the value key equals to 128.

B. Challenge II: How to Combine the Leakage Information from Multiple Leak Sites

Real-world software can have various side-channel vulnerabilities. Those vulnerabilities may spread in the whole program. An adversary may exploit more than one side-channel vulnerabilities to gain more information [4], [5]. In order to precisely quantify the total information leakage, we need to know the relation of those leakage sites.

```

1 unsigned char k1, k2, k3;
2 ...
3 t1 = T[k1 % 8];           // Leakage 1
4 t2 = T[k2 % 8];           // Leakage 2
5 t3 = T[k3 % 8];           // Leakage 3
6 bool flag = foo(t1, t2, t3);
7 if(flag) {
8     if(k1 > 128)           // Leakage 4
9         A();
10 }
```

```

11 else {
12     if(k2 > 128)           // Leakage 5
13         B();
14 }
15 if(k1 + k2 + k3 > 128) // Leakage 6
16     C();

```

Listing 2: Multiple leakages

Consider the running example in 2, in which k_1 , k_2 and k_3 are the sensitive key. The code has six different leakage. Leakage 1, 2, 3 are the secret-dependent data accesses and leakage 4, 5, 6 are the secret-dependent control-flow transfers. The attacker can infer the last three digits of k_1 , k_2 , k_3 from leakage 1, 2, 3. So those leakages are independent. For leakage 1, 4, 6, however, we have no idea about the total information leakage.

Suppose one program has two side-channel vulnerabilities A and B, which can leak L_A and L_B bits respectively according to the definition 1. Depending on the relation between A and B, the total leaked information L_{Total} will be:

1) *Independent Leakages*: If A and B are independent leakages, the total information leakage will be:

$$L_{total} = L_A + L_B$$

2) *Dependent Leakages*: If A and B are dependent leakages, the total information leakage will be:

$$\max\{L_A, L_B\} \leq L_{total} < L_A + L_B$$

3) *Mutual Exclusive Leakages*: If A and B are mutual exclusive leakages, then only A or B can be observed for one fixed input. The total information leakage will be L_A or L_B .

According to above definition, leakage 1, 2, 3 are independent leakages. Leakage 4, 5 are mutual exclusive leakages. For real-world applications, it is hard to estimate the total leaked information for the following reasons. First, the real-world applications have more than thousands of lines of code. One leakage site leaks the temporary value. But the value contains some information about the original buffer. It is hard to know how the sensitive value affects the temporary value. Second, some leakages sites may be dependent. The occurrence of the first affects the occurrence of the second sites. We can't simply add them up. Third, leakage sites are in the different blocks of the control-flow graph, which means that only one of the two leakages site may be executed during the execution.

Our Solution to Challenge II: Given a program P with k as the sensitive input, we use k_i to denote the sensitive information, where i is the index of the byte in the original buffer. We can represent each temporal values with a formula. There are two types of values in the formula: the concrete value and the symbolic value. We use the runtime information to simplify the formula. In other words, we only use symbolic values to represent the sensitive input. For other values that are independent from the sensitive input, we use the concrete value from the runtime information.

After that, we model each leakage sites as a math formulas. The attacker can retrieve the sensitive information by observing the different patterns in control-flows and data access when the program process different sensitive information. We refer them as the secret-dependent control flow and secret-dependent data access accordingly. For secret-dependent control transfers, we model the leakage using the path conditions that cause the control transfer. For secret-dependent memory accesses, we use a symbolic formula $F(\vec{K})$ to represent the memory address and check if different sensitive inputs can lead to different memory accesses. As long as we model each leakage with a formula. We can regard multiple leakages as the conjunction of those formulas.

C. Challenge III: Scalability and Performance

After we transfer each potential leaks sites into formula. We can group several formulas together to estimate the total information leakage. One naive way is to use the Monte Carlo sampling estimate the number of input keys. With the definition 1, we can estimate the total information leakage.

However, some pre-experiments show that above approach suffers from the unberable cost, which impede its usage to detect and quantify side-channel leakages in real-world applications. We systematically analyze the performance bottlenecks of the whole process. In general, the performance suffers from the two following reasons.

- Symbolic Execution (Challenge III(a))
- The Monte Carlo Sampling (Challenge III(b))

1) *Symbolic Execution*: Symbolic execution interprets each instruction and update the memory cells and registers with a formula that captured the semantics of the execution. Unfortunately, the number of machine instructions are huge and the semantics of each instruction is complex. For example, the Intel Developer Manual [9] introduces more than 1000 different X86 instructions. It is tedious to manually implement the rules for every instructions.

Therefore, existing binary analysis tools [10], [11] will translate machine instructions into intermediate languages (IR). The IR typically has fewer instructions compared to the original machine instructions. The IR layer designs, which significantly simplify the implementations, also introduce significant overhead as well [12].

Our Solution to Challenge III(a): We adopt the similar approach from [12] and implement the symbolic execution directly on the top X86 instructions.

2) *Monte Carlo Sampling*: For an application with m bytes secret, there are total 2^{8m} possible inputs. Of the 2^{8m} possible inputs, we want the number of inputs which satisfy those formulas. Then we can use the definition /refdef to calculate the information leakage.

A Monte Carlo method for approximating the number of $|K_o|$ is to pick up M random values and check how many of them satisfy those constrains. If l values satisfy those constrains, then the approximate result is $\frac{l * 2^{8m}}{M}$.

However, the number of satisfying values could be exponentially small. Consider the formula $F = k_1 = 1 \wedge k_2 =$

$2 \wedge k_3 = 3 \wedge k_4 = 4$, k_1 , k_2 , k_3 and k_4 each represent one byte in the original sensitive input, there is only one possible solution of 2^{32} possible values, which requires exponentially many samples to get a tight bound. The naive Monte Carlo Method also suffers from the curse of dimensionality. For example, the libjpeg libraries can transfer the image from one format into another format. One image could be 1kMB. If we take each byte in the original buffer as symbols, the formula can have at most 1024 symbols.

Our Solution to Challenge III(b): We adopt the Markov Chain Monte Carlo to estimate the number of possible input that satisfies the logic formula groups. The key idea is that we have one group of input that satisfies the logic formula constrains. We will introduce the method in the following sections.

VI. DESIGN

In this section, we describe the design of CleverHans by focusing on how our design solves the three challenges discussed in the previous section.

A. Workflow

The shortcomings of the existing work inspire us to design a new tool to detect and quantify information leakage vulnerabilities in binaries. The tool has three steps. First, we run the target program with the concrete input (sensitive information) under the dynamic binary instrumentation (DBI) frameworks to collect the execution traces. After that, we run the symbolic execution to capture the fined-grained semantic information of each secret-dependent control-flow transfers and data-accesses. Finally we run the Markov Chain Monte Carlo (MCMC) to estimate the information leakage.

- 1) *Execution trace generation.* The design goal of CleverHans is to estimate the information leakage as precisely as possible. Therefore, we sacrifice the soundness for the precision in terms of program analysis. Previous works [2], [3] have demonstrated the effectiveness of the dynamic program analysis. We run the target binary under the dynamic binary instrumentation (DBI) to record the execution trace and the runtime information.
- 2) *Instruction level symbolic execution.* We model the attacker's observation about the side-channel vulnerabilities with math formula. Each formula capture the fined-grained information between the input secrets and the leakage site. For the consideration of precision and performance, we remove the intermediate language(IR) layer of the symbolic execution. Also, the engine only symbolically execute the instruction that might be affected the input key. The above design significantly reduces the overhead of the symbolic execution, which make the tool scales to real-world programs.
- 3) *Leakage estimation.* We transfer the information leakage quantification problem into the problem of counting the number of assignments that satisfy the formulas which models the observations from attacker. We propose a markov monte carlo method to estimate the number

of satisfied solutions. With the help of Chebyshev's Inequality, we also give the an error estimate with the probability.

B. Trace Logging

The trace information can be logged via some emulators (e.g., QEMU) or dynamic binary instrumentation tools (DBI). We run a program with the concrete input under the DBI to record the execution trace. The trace data has the following information:

- Each instruction mnemonics and its memory address.
- The operands of each instruction and their concrete values during the runtime.
- The value of eflags register.
- The memory address and the length of the sensitive information. Most software developers stores sensitive information in an array, a variable or a buffer, which means that those data is stored in a contiguous area in the memory. We use the symbol information in the binary to track the address in the memory.

C. Instruction Level Symbolic Execution

The main purpose of the step is to generate constraints of the input sensitive information from the execution trace. If we give the target program a new input which is different from the origin input that was used to generate the execution trace but still satisfies those constraints, the new execution trace still have the same control flow and data access patterns.

The tool runs the symbolic execution on the top of the execution traces. At the beginning of the symbolic execution, the tool creates fresh symbols for each byte in the sensitive buffer. For other data in the register or memory at the beginning, we use concrete values from the runtime information collected in the previous step. During the symbolic execution for each instruction, the tool updates every variable in the memory and registers with a math formula. The formula is made up with concrete values and the input key as the symbols accumulated through the symbolic execution. For each formula, the tool will check weather it can be reduced into a concrete values (e.g., $k_1 + 12 - k_1 = 12$). If so, the tool will only use the concrete values in the following symbolic execution.

1) *Verification and Optimization:* We run the symbolic execution(SE) on the top of x86 instructions. In other words, we dont rely on any intermediate languages to simplify the implemetation of symbolic execution. While the implementation itself has a lot of benefits (Better performance, accurate memory model), we need to implement the symbolic execution rules for each X86 instruction. However, due to the complexity of X86, it is inevitable to make mistakes. Therefore, we verify the correctness of the SE engine during the execution. The tool will collect the runtime information (Register values, memory values) and compare them with the formula generated from the symbolic execution. Whenever the tool finishes the symbolic execution of each instruction, the tool will compare the formula for each symbol and its actual value. If the two values don't match, we check the code and fix the error. Also,

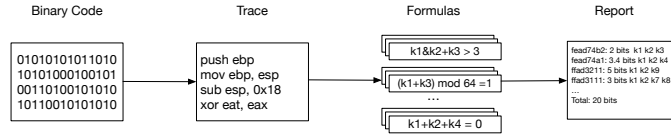


Fig. 2: The workflow of CleverHans. **FIXME:** redraw the figure. i will give you the instruction.

if the formula doesn't contain any symbols, the tool will use the concrete value instead of symbolic execution.

2) *Secret-dependent control-flows*: An adversary can infer sensitive information from secret dependent control-flows. There are two kinds of control-transfer instructions: the unconditional control-transfer instructions and the conditional transfer instructions. The unconditional instructions, like CALL, JUMP, RET transfer control from one code segment location to another. Since the transfer is independent from the input sensitive information, an attacker was not able to infer any sensitive information from the control-flow. So the unconditional control-transfer doesn't leak any information based on our threat model. During the symbolic execution, we just update the register information and memory cells with new formulas accordingly.

The conditional control-flow transfer instructions, like conditional jumps, depending on CPU states, may or may not transfer control flows. For conditional jumps, the CPU will test if certain condition flag (e.g., CF = 0, ZF = 1) is met and jump to the certain branches respectively. The symbolic engine will compute the flag and represent the flag in a symbol formula. Because we are running on a symbolic execution on a execution trace, we know which branch is executed. If a conditional jump uses the CPU status flag, we will generate the constraint accordingly.

```

1 804841c; movzx edx, byte ptr [ebp+var_D];    var_D = k1;  edx = zeroext(k1)
2 8048420; movzx eax, byte ptr [ebp+var_C];    var_C = k2;  eax = zeroext(k2)
3 8048424; add eax, edx;                      eax = zeroext(k1) + zeroext(k2); edx = zeroext(k1)
4 8048426; mov byte ptr [ebp+var_B], al;      var_B = k1 + k2
5 8048429; movzx edx, byte ptr [ebp+var_C];    var_C = k2;  edx = zeroext(k2)
6 804842d; movzx eax, byte ptr [ebp+var_A];    var_A = k3;  eax = zeroext(k3)
7 8048431; sub edx, eax;                      eax = zeroext(k3);  edx = zeroext(k2) - zeroext(k3)
8 8048433; sub eax, edx;                      eax = zeroext(k2) - zeroext(k3)
9 8048435; mov byte ptr [ebp+var_9], al;      var_9 = k2 - k3
10 8048438; movzx eax, byte ptr [ebp+var_B];    eax = k1 + k2
11 804843c; cmp al, byte ptr [ebp+var_9];      CF = (k1 + k2) > (k2 - k3)
12 804843f; jnb 0x8048451;                    Formula: k1 ≤ k3
13 8048441; sub esp, 0xc;                     esp = C1 - 0xc

```

Fig. 3: The workflow of CleverHans

For examples,

```

...
0x0000e781    add dword [local_14h], 1
0x0000e785    cmp dword [local_14h], 4
0x0000e789    jne 0xe7df
0x0000e78b    mov dword [local_14h], 0
...

```

At the beginning of the instruction segment, the value at the address of local14h can be written as $F(\vec{K})$. At the address e785, the value will be updated with $F(\vec{K}) + 1$. Then the code compares the value with 4 and use the result as a conditional jump. Based on the result, we can have the following formula:

$$F(\vec{K}) + 1 = 4$$

The formula, together with the memory address (0xe789) is store as a *formula tuple* (address, formula). Each formula tuple represents one leakage site.

3) *Secret-dependent data access*: Like input-dependent control-flow transfers, an adversary can also infer sensitive information from the data access pattern as well. We try to find this kind of leakages by checking every memory operand of the instruction. We generate the memory addressing formulas. As discussed before, every symbols in the formula is the input key. If the formula doesn't contain any symbols, the memory access is independent from the input sensitive information and won't leak any sensitive information according to our threat model. Otherwise, we will generate the constraint for the memory addressing. We model the memory address with a symbolic formula $F(\vec{K})$. Because we also have the concrete value of the memory address Addr1. Inspired by the work from [2], the formula can be written as:

$$F(\vec{K}) \gg L = \text{Addr1} \gg L$$

The L represents the minimum memory address granularity that an attacker can observe. For example, Flush and Reload can distinguish between different cache lines, which means the value of L is 6.

D. Information Leakage Estimation

In this section, we present the algorithm to calculate the information leakage based on the definition 1.

1) *Problem Statement*: From the above step VI-C, we can generate the constraint for each unique leakage site on the execution trace. The only variables in those constraints are the sensitive data represented with k_1, k_2, \dots, k_n . Suppose the address of the leakage site is i , we use C_i to denote the constraint. For multiple leakage sites, as each leakage site has one constraint, we can use the conjunction of those constraints to represent those leakage sites.

According to the definition 1, to calculate the amount of leaked information, the key is to calculate $\frac{|K|}{|K^o|}$. $|K^o|$ represents the set that contains every input keys that satisfy the constraint. As the cardinality of K is known, the key problem is to know the cardinality of K^o . Suppose an attacker can observe n leakage sites, and each leakage site has the following constraints: C_1, C_2, \dots, C_n respectively. The total leakage has the constraint $F_{c1c2\dots cn} = C_1 \wedge C_2 \wedge C_3 \wedge \dots \wedge C_n$. The problem of estimating the total leaked information can be

transferred into counting the number of different solutions that satisfies the constraint $F_{c_1c_2...c_n}$.

A native method for approximating the result is to pick k elements from K and check how many of them are also contained in K^o . If q elements are also in K^o . In expectation, we can use $\frac{k}{q}$ to approximate the value of $\frac{|K|}{|K^o|}$.

However, as the discussion in V-C2, the above sampling method will typically fail in practice due to the following two problems:

- 1) The curse of dimensionality. $F_{c_1...c_n}$ is the conjunction of many constraints. Therefore, the input variables of each constraints will also be the input variables of the $F_{c_1...c_n}$. The sampling method will fail as n increases. For example, when n equals to 2, the whole search space is a 256^2 cube. If we want the distance between each points equals to 1, we need 256^2 points. When n equals to 10, we need 256^{10} points if we still want the distance between each points equals to 1. As a result, the error of the result based on sampling method will increase exponentially with adding dimensions.
- 2) The number of satisfying assignments could be exponentially small. According to Chernoff bound, we need exponentially many samples to get a tight bound. On an extreme situation, if the constraint only has one unique satisfying solution, the simple Monte Carlo method can't find the satisfying assignment after sampling many points.

In regard to the above problems, we present our methods. First, we split $F_{c_1...c_n}$ into several independent constraint groups. After that, we run random-walk based sampling method on each constraint.

2) *Maximum Independent Partition*: For a constraint C_i , we define a function S , mapping the constraint into a set of input symbols. For example, $S(k_1 + k_2 > 128) = \{k_1, k_2\}$.

Definition VI.1. Given two constraints C_m and C_n , we call them independent iff

$$S(C_m) \cap S(C_n) = \emptyset$$

Based on the definition, the first step is to split the constraint $F_{c_1c_2...c_n}$ into several independent constraints. There are many partitions. For our project, we are interested in the following one.

Definition VI.2. For the constraint $F_{c_1c_2...c_n}$ and the constraint group $C_{g1}, C_{g2}, ..., C_{gm}$, we call $C_{g1}, C_{g2}, ..., C_{gm}$ is the maximum independent partition of $F_{c_1c_2...c_n}$ iff

- 1) $C_{g1} \wedge C_{g2} \wedge ... \wedge C_{gm} = F_{c_1c_2...c_n}$
- 2) $\forall i, j \in \{1, 2, 3, ..., m\}$ and $i \neq j$, $S(C_i) \cap S(C_j) = \emptyset$
- 3) For any other partitions $C_{g1}, C_{g2}, ..., C_{gn}$ satisfy 1) and 2), $m \geq n$

The reason we want a good partition of the constraints is that we want to reduce the dimensions. Consider the example in the previous section,

$$F : k_1 = 1 \wedge k_2 = 2 \wedge k_3 = 3 \wedge k_4 = 4$$

The good partition of F would be

$$C_{g1} : k_1 = 1 \quad C_{g2} : k_2 = 2 \quad C_{g3} : k_3 = 3 \quad C_{g4} : k_4 = 4$$

So instead of sampling in the four dimension space, we can sample each constraint in the one dimension space and combine them together.

Theorem VI.1. $C_{g1}, C_{g2}, ..., C_{gm}$ is a maximum independent partition of $F_{c_1c_2...c_n}$. K_c is the input set that satisfies constrain c . We can have the following equation in regard to the size of K_c

$$|K_{F_{c_1c_2...c_n}}| = |K_{C_{g1}}| * |K_{C_{g2}}| * ... * |K_{C_{gm}}|$$

With the VI.1, we can transfer the problem of counting the number of solutions in a large constraint with high dimensions into counting solutions of several small constraints.

3) *Motivation*: The adversary, however, can observe the memory access pattern of the software. If the attacker can observe one leakage site, the attack is modeled as one formula. If he observes multiple information leakage, the attack is modeled as the conjoint of those formulas. Calculating the total information leakage can be reduced to the problem of approximating the number of solutions. This is a #P problem.

One intuition way is to use the Monte Carlo method. However, as the discussion in the previous section V-C2, the number of satisfying keys could be exponentially small. Imagine an attacker who can recover one unique key after the attack, in such case, the conjoint of formula $F = (f(k))$ only has one solution. Also, the target program may have multiple inputs. When the number of input symbols arises, the simple Monte Carlo may suffer from the curse of dimensionality, which means the accuracy of the result drops dramatically when the number of input symbols increases.

We adopt the Markov Chain Monte Carlo (MCMC) to estimate the number of satisfying solutions. The basic idea is to construct a Markov Chain that has the desired distribution.

Definition: Starting from here, we present the formal definition of the MCMC algorithm. First, we introduce the definition of the concept of the algorithm.

Theorem VI.2. If $F = A_1, A_2, ..., A_n$ is a finite collection of closed sets then $\cup_i^n A_i$ is a closed set.

4) *MCMC Algorithm*: In the section, we present the detail of MCMC algorithm for counting the cardinality of K^o . We will first present the description of the MCMC and then explain the algorithm with a concrete example.

- 1) Start with the concrete input $\vec{k} = (k_0, k_1, k_2, ..., k_n)$, \vec{k} should be the on valid solution which satisfies the constraint $F = C_1(\vec{k}) \wedge C_2(\vec{k}) \wedge C_3(\vec{k}) \wedge ... \wedge C_m(\vec{k})$
- 2) for $t = 1, 2, 3, ..., N$

VII. IMPLEMENTATION

We implement the CleverHans with **FIXME**: 12K, be precise lines of code in C++11. It has three components, Intel Pin tool that can collect the execution trace, the instruction-level symbolic execution engine and the backend that can estimate the information leakage. The tool can also report the memory

address of the leakage site. To assist the developers fix the bugs more easily, we also have several Python scripts that can report the leakage location in the source code with the debug information and the symbol information. A sample report can be found at the appendix.

Component	Lines of code (LOC)
Trace Logging	101 LOC of C++
Symbolic Execution	11547 LoC of C++
Monte Carlo Sampling	603 LOC of C++
Others	FIXME: xxx LOC of Python
Total	FIXME: xxx LOC

TABLE III: CleverHans’s main components and their LoC

Our current implementation can support part of the Intel 32-bit instructions that are important in finding memory-based side-channels vulnerabilities, which include bitwise operations, control transfer, data movement and logic instructions. For other instructions the current implementation doesn’t support, the tool will use the concrete values to update the registers and memory cell. Therefore, the tool may miss some leakages but won’t give us any new false positives with the implementation.

VIII. EVALUATION

We evaluate CleverHans with the real-world crypto libraries and non-crypto. For crypto libraries, we choose OpenSSL and mbedTLS, two most commonly used crypto libraries in today’s software. For OpenSSL, we compile it with the option *no-asm* to disable the assembly language routines. For non-crypto libraries, we study the libjpeg, a commonly used lossy library implemented by the Independent JPEG Group. We build the source code into a 32 bit ELF binary with the GCC 8.0 on Ubuntu 14.04. Although our tool can work on stripped binaries, we use the symbol information to trace back leakage sites into the source code. We use Intel Pin version 3.7 to record the execution trace.

During the evaluate, we are interested in the following aspects:

- 1) **Finding Leakage Sites.** How effective is CleverHans in identifying the memory-based side-channels vulnerabilities? How much performance overhead does the tool need to find the leakage?
- 2) **Leakage Sites Quantification.** Can CleverHans precisely report the information leakage? Is the number of leaked bits reported by CleverHans useful to justify the sensitive level of the side-channel vulnerability?
- 3) **Exsiting Leakage Sites.** Recent work has report a number of leakage sites in open source crypto libraries. But the crypto library authors don’t fix them for some reasons. We use CleverHans to study those vulnerabilities and try to explain the reason why the author fix or not fix the side channel vulnerabilities.

A. Finding Leakage Sites

B. Leakage Sites Quantifications

C. Exsiting Leakage Sites

IX. DISCUSSION

A. Binary code vs source code

Many existing works find side-channels vulnerabilities from source code level or intermediate languages (e.g., LLVM IR). Those approaches will have the following questions. First, many compilers can translate the operator into branches. For example, the GCC compiler will translate the ! operator into conditional branches. If the branch is secret-dependent, the attacker could learn some sensitive information. But those source-based methods will fail to detect those vulnerabilities. Second, some if else in the source code can be converted into single conditional instructions (e.g., *cmov*). Source-based methods will still regard them as potential leakages, which will introduce false positives.

B. X86 vs Intermediate Languages

The tool takes in an unmodified binary and the marked sensitive information as the input. The tool will first start with logging the execution trace. Then the tool will symbolizes the sensitive information into multiple symbols and start with the symbolic execution. During the symbolic execution, the tool will find any potentials leakage sites as well as the path constraints. For each leakage site, whether it is the tool will also generate the constraint. Then, the tool will split the constraints into multiple group. After that, the tool will run monte carlo sampling to estimate the total information leakages.

X. RELATED WORK

A. Side-channel vulnerabilities detection

There are a large number of works on side-channel vulnerability detections in recent years. CacheAudit [1] uses abstract domains to compute the overapproximation of cache-based side-channel information leakage upper bound. However, due to over approximation, the leakage given by CacheAudit can indicate the program is side-channel free if the program has zero leakage. However, it is less useful to judge the sensitive level of the side-channel leakage based on the leakage provided by CacheAudit. CacheS [13] improves the work of CacheAudit by proposing the novel abstract domains, which only precisely track secret related code. Like CacheAudit, CacheS can’t provide the information to indicate the sensitive level of side-channel vulnerabilities.

The dynamic approach, usually comes with taint analysis and symbolic execution, can perform a very precise analysis. CacheD [2] takes a concrete program execution trace and run the symbolic execution on the top of the trace to get the formula of each memory address. During the symbolic execution, every value except the sensitive key uses the concrete value. Therefore, CacheD is quite precise in term of false positives. We adopted the similar idea to model the secret-dependent memory accesses. DATA [3] detects address-based

Program	Implementation	Start Function	Key Size (bits)	CF	DA	Total Leakage	Number of Instructions	Process Time (s)
AES	OpenSSL 0.9.7	AES_set_encrypt_key						
DES	OpenSSL 0.9.7	DES_string_to_key						
RSA	OpenSSL 0.9.7							
AES	OpenSSL 1.1.1							
DES	OpenSSL 1.1.1							
RSA	OpenSSL 1.1.1							
AES	MBED TLS 2.5							
DES	MBED TLS 2.5							
RSA	MBED TLS 2.5							
AES	MBED TLS 2.15							
DES	MBED TLS 2.15							

side-channel vulnerabilities by comparing execution traces under various tests.

B. Quantification

Quantitative Information Flow (QIF) aims at providing an information leakage estimate for the sensitive information given the public output. If zero bits of the information is leaked, the program is called non-interference. McCamant [14] quantify the information leaked as the network flow capacity. Phan [15] propose symbolic QIF. The goal of their work is to ensure the program is non-interference. They adopt an over approximation way of estimating the total information leakage and their method doesn't work for secret-dependent memory access side-channels. CHALICE [16] quantifies the information leakage based on Cache miss and Cache hit behaviours. But CHALICE can only work on PICA code, which limits its usage to X86 binary. Due to complexity of the Cache model and Cartesian Bounding they use, CHALICE can only scale to small programs, which limits its usage in real-world applications.

XI. CONCLUSION

In this paper, we present CleverHans for identifying and quantifying memory-based side-channel leakages. We show that CleverHans is effective at finding and quantifying the side-channel leakages. With the new definition of information leakage that imitates an real side-channel attacker, the number of leaked bits is useful to justify the understand the sensitive level of side-channel vulnerabilities. The evaluation results confirm our design goal and show CleverHans is useful to estimate the amount of leaked information in real-world applications.

REFERENCES

- [1] G. Doychev, D. Feld, B. Kopf, L. Mauborgne, and J. Reineke, "Cacheaudit: A tool for the static analysis of cache side channels," in *Presented as part of the 22nd USENIX Security Symposium (USENIX Security 13)*. Washington, D.C.: USENIX, 2013, pp. 431–446. [Online]. Available: <https://www.usenix.org/conference/usenixsecurity13/technical-sessions/paper/doychev>
- [2] S. Wang, P. Wang, X. Liu, D. Zhang, and D. Wu, "CacheD: Identifying cache-based timing channels in production software," in *26th USENIX Security Symposium (USENIX Security 17)*. Vancouver, BC: USENIX Association, 2017, pp. 235–252. [Online]. Available: <https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/wang-shuai>
- [3] S. Weiser, A. Zankl, R. Spreitzer, K. Miller, S. Mangard, and G. Sigl, "DATA – differential address trace analysis: Finding address-based side-channels in binaries," in *27th USENIX Security Symposium (USENIX Security 18)*. Baltimore, MD: USENIX Association, 2018, pp. 603–620. [Online]. Available: <https://www.usenix.org/conference/usenixsecurity18/presentation/weiser>
- [4] D. Gruss, R. Spreitzer, and S. Mangard, "Cache template attacks: Automating attacks on inclusive last-level caches," in *24th USENIX Security Symposium (USENIX Security 15)*. Washington, D.C.: USENIX Association, 2015, pp. 897–912. [Online]. Available: <https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/gruss>
- [5] Y. Xu, W. Cui, and M. Peinado, "Controlled-channel attacks: Deterministic side channels for untrusted operating systems," in *2015 IEEE Symposium on Security and Privacy*, May 2015, pp. 640–656.
- [6] T. Hornby, "Side-channel attacks on everyday applications: distinguishing inputs with FLUSH+RELOAD. black hat usa (2016)," 2011.
- [7] G. Smith, "On the foundations of quantitative information flow," in *Foundations of Software Science and Computational Structures*, L. de Alfaro, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 288–302.
- [8] B. Kopf, L. Mauborgne, and M. Ochoa, "Automatic quantification of cache side-channels," in *Computer Aided Verification*, P. Madhusudan and S. A. Seshia, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 564–580.
- [9] Intel Corporation, *Intel® 64 and IA-32 Architectures Software Developer's Manual*, December 2009, no. 253669-033US.
- [10] Y. Shoshitaishvili, R. Wang, C. Salls, N. Stephens, M. Polino, A. Dutcher, J. Grosen, S. Feng, C. Hauser, C. Kruegel, and G. Vigna, "SoK: (State of) The Art of War: Offensive Techniques in Binary Analysis," in *IEEE Symposium on Security and Privacy*, 2016.
- [11] D. Brumley, I. Jager, T. Avgerinos, and E. J. Schwartz, "Bap: A binary analysis platform," in *Computer Aided Verification*, G. Gopalakrishnan and S. Qadeer, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 463–469.
- [12] I. Yun, S. Lee, M. Xu, Y. Jang, and T. Kim, "QSYM: A practical concolic execution engine tailored for hybrid fuzzing," in *27th USENIX Security Symposium (USENIX Security 18)*. Baltimore, MD: USENIX Association, 2018, pp. 745–761. [Online]. Available: <https://www.usenix.org/conference/usenixsecurity18/presentation/yun>
- [13] "Identifying cache-based side channels through secret-augmented abstract interpretation," in *28th USENIX Security Symposium (USENIX Security 19)*. Santa Clara, CA: USENIX Association, 2019. [Online]. Available: <https://www.usenix.org/conference/usenixsecurity19/presentation/wang-shuai>
- [14] S. McCamant and M. D. Ernst, "Quantitative information flow as network flow capacity," in *PLDI 2008: Proceedings of the ACM SIGPLAN 2008 Conference on Programming Language Design and Implementation*, Tucson, AZ, USA, Jun. 2008, pp. 193–205.
- [15] Q.-S. Phan, P. Malacaria, O. Tkachuk, and C. S. Pășăreanu, "Symbolic quantitative information flow," *SIGSOFT Softw. Eng. Notes*, vol. 37, no. 6, pp. 1–5, Nov. 2012. [Online]. Available: <http://doi.acm.org/10.1145/2382756.2382791>
- [16] S. Chattopadhyay, M. Beck, A. Rezzine, and A. Zeller, "Quantifying the information leak in cache attacks via symbolic execution," in *Proceedings of the 15th ACM-IEEE International Conference on Formal Methods and Models for System Design*, ser. MEMOCODE '17. New York, NY, USA: ACM, 2017, pp. 25–35. [Online]. Available: <http://doi.acm.org/10.1145/3127041.3127044>