Comparative Analysis of Leakage Tools on Scalable Case Studies

Fabrizio Biondi(✉), Axel Legay, and Jean Quilbeuf(✉)

Inria, Rennes, France
{fabrizio.biondi,jean.quilbeuf}@inria.fr

Abstract. Quantitative security techniques have been proven effective to measure the security of systems against various types of attackers. However, such techniques are often tested against small-scale academic examples.

In this paper we analyze two scalable, real life privacy case studies: the privacy of the energy consumption data of the users of a smart grid network and the secrecy of the voters' voting preferences with different types of voting protocols.

We contribute a new trace analysis algorithm for leakage calculation in QUAIL. We analyze both case studies with three state-of-the-art information leakage computation tools: LeakWatch, Moped-QLeak, and our tool QUAIL equipped with the new algorithm. We highlight the relative advantages and drawbacks of the tools and compare their usability and effectiveness in analyzing the case studies.

1 Introduction

The protection of privacy and data security is one of the main concerns of computer science. Security often falls down to the impossibility for an attacker to obtain a given secret value. Such an impossibility can be defined by non-interference [18]. However this definition rejects any program which publishes any variable whose value depends on the secret. For instance, publishing the results of an election when each individual vote is secret breaks non-interference. Such a yes/no approach does not consider that an attacker may have a partial information about a secret.

Information-theoretical techniques have the advantage of considering the secret not as an atomic object but as a known number of secret bits, allowing the definition of measures of effectiveness of an attack based on the amount of secret bits that the attack compromises. The amount of secret bits that are compromised by an attack are known as *information leakage*. Leakage depends on the information about the secret known to the attacker before the attack, known as *prior information* and usually modeled as a *prior probability distribution* over the values of the secret. This approach dates back to Denning [16]. Different information leakage measures have been introduced, including Shannon leakage [19], min-entropy leakage [30] and the g-leakage [1], encoding different security properties of the system. All the tools we compare in this work can

B. Fischer and J. Geldenhuys (Eds.): SPIN 2015, LNCS 9232, pp. 263–281, 2015.
DOI: 10.1007/978-3-319-23404-5_17

compute both Shannon and min-entropy leakage with no significant difference in computation time. We compare them on the computation of Shannon leakage, but we expect no significant difference if the tools were to be compared on the computation of min-entropy leakage.

Among the results in the field, Köpf et al. studied leakage of side-channel attacks [2,21], while Boreale has defined leakage for process calculi [6] and characterized the best attack strategy of an adaptive attacker [9].

In this work we compare the three tools that compose the state of the art in Shannon leakage computation: QUAIL [5] equipped with a new trace analysis algorithm, LeakWatch [13], and Moped-QLeak [11].

QUAIL is a recent but already well established tool for precise and exact information leakage computation, and later tools by multiple authors have used it as comparison [13,25]. Nonetheless, QUAIL needs to produce a full Markov chain model of the system-attacker scenario to produce a meaningful result.

LeakWatch is the most recent of a family of tools for statistical approximation of information leakage developed by Chothia et al. [12,14]. LeakWatch analyzes Java code, requiring the programmer to annotate the code of the system with secret and observable values, then simulates the system repeatedly using the Java Virtual Machine and estimates the correlation between the secret and observable values. LeakWatch follows a different perspective than QUAIL and Moped-QLeak, since LeakWatch computes an approximated result, contrarily to QUAIL's arbitrary precision and Moped-QLeak's fixed double precision. LeakWatch's approximation can be improved at the cost of running more simulations, which is time expensive.

Moped-QLeak [11] uses the Moped tool [17] to compute a symbolic summary of the program under analysis as an Algebraic Decision Diagram (ADD), and then computes the leakage using the ADD representation. The symbolic approach is very efficient when the program can be represented in a compact way using ADDs, and in these cases Moped-QLeak is significantly faster than the other tools.

The first contribution of this paper is a new algorithm for precise information leakage computation, which is able to compute information leakage following the same Markovian semantics we introduced previously [3, Sect. 4] by performing a depth-first search analyzing the execution traces of the system. We implemented this algorithm in QUAIL, allowing it to compute leakage without having to build the full Markov chain model of the system.

As a second contribution, we provide two scalable case studies for the benchmarking of quantitative information leakage tools. Both case studies arise from real-life privacy problems. The case studies are anonymity of user data in Smart Grids and privacy comparison in voting protocols.

Smart Grids are in the family of interconnected objects and have received a growing interest over the last years. Our case study is based on a real system deployed at fortiss[1] labs [22]. In our case study, we focus on the negotiation between a set of *prosumers* and an *aggregator*. The prosumers (PROducer conSUMERS) consume, store and produce energy. To stabilize the grid, the

[1] http://fortiss.org.

prosumers negotiate with the aggregator how much energy they will exchange with the grid for the next period of time. This exchange might expose the consumption of one of the prosumers, and, in turn, allow a potential attacker to deduce that a house is empty or that a factory has increased its production. In that example, the difficulty is to decide not only whether the exact information can be deduced or not, but also how well an attack can approximate it. Measuring the leakage indicates how much of the secret is unveiled through the negotiation phase. We show that increasing the number of prosumers also increases security.

In the voting protocols comparison case study, we compare two different voting protocols: the *Single Preference*, where each voter expresses a single vote for his favorite candidate, and the *Preference Ranking*, where each voter ranks all candidates from his most to his least favorite. In both cases there are multiple voters and candidates, and the secret is the preference of each voter. Both protocols have a large number of possible secrets and outputs, so they become cumbersome to analyze even with a small number of voters and candidates.

We compare the tools on their computation time, precision of the answer returned, scalability and usability. Since no tool works strictly better than the others in all category, we determine the problem classes that are better suited to be analyzed by each tool.

2 Background: Information Leakage

The information leakage of a program is a measure quantifying how much information an attacker infers about the program's secret by observing the program's output. We assume that the attacker has access to the program's source code, unlimited computational power, and some prior information about the secret (e.g. the bit size of the secret). Leakage corresponds to the reduction in the attacker's uncertainty about the secret.

Let $\mathtt{h}$ be a random variable with values in a domain $D(\mathtt{h})$ representing the value of the secret and $\mathtt{o}$ be a random variable with values in a domain $D(\mathtt{o})$ modeling the value of the output. The information the attacker has on the secret is modeled by a discrete probability distribution, i.e. for a discrete random variable X a function $\pi : D(X) \to [0, 1]$ such that $\sum_{x \in D(X)} \pi(x) = 1$. The information that the attacker has on the secret before the attack is modeled by the *prior distribution* $\pi(h)$ while the information the attacker has after observing the output is modeled by the *posterior distribution* $\pi(h|o)$. We consider the prior distribution as given, since it is part of the model of the attacker. Let U be an uncertainty measure defined on probability distributions, including Shannon entropy, min-entropy, and g-vulnerability. Computing leakage for the measure U reduces to computing the prior and posterior distributions and applying the formula

$$Leakage_U = U(\pi(h)) - U(\pi(h|o)) \tag{1}$$

$$= U(\pi(h)) - \sum_{\bar{o} \in D(o)} \pi(o = \bar{o}) U(\pi(h|o = \bar{o})) \tag{2}$$

In this work we want to compute Shannon leakage, and thus we use Shannon entropy as the measure of uncertainty: $U(\pi(x)) = \sum_{x \in D(X)} \pi(x) \log_2 \pi(x)$.

3 Quantitative Information Leakage Tools

We introduce the quantitative information leakage computation tools that will be tested on the case studies.

3.1 QUAIL

QUAIL [5] computes Shannon and min-entropy leakage of a program written in an imperative WHILE language. The language allows the user to program naturally with constants, arrays, and loops, which is syntactic sugar for QUAIL's if-goto Markovian semantics. Given the prior information of the attacker, QUAIL represents the program as a Markov chain, and computes the information leakage from the Markov chain with an arbitrary number of precision digits.

Syntax. We present the syntax of the QUAIL imperative language we use to model programs. We distinguish the variables in *public* and *private* variables according to their level of abstraction: public variables have precise values, while private variables have sets of possible values. The observable variable `o` is public, while the secret variable `h` is private. Let `v` range over names of variables and x range over reals from $[0; 1]$. Let L (resp. H) be a set of assignments of values to public variables (resp. assignments of sets of values to private variables).

Let *label*, l_0 and l_1 denote any program point and `f` (`g`) pure arithmetic (Boolean) expressions. Assume a standard set of expressions and the following statements:

stmt ::= `public int`n `v :=` k | `private int`n `v` | `v := f(`L`)` | `v := rand` x |
`skip` | `goto` *label* | `return` | `if g(`L, H`) then goto` l_a
`else goto` l_b

The first statement declares a public variable `v` of size n bits with a given value k, while the second statement similarly declares a private variable `h` of size n bits with allowed values ranging from 0 to $2^n - 1$. We assume a standard type system to verify that values of n-bit variables do not exceed $2^n - 1$. The third statement assigns to a public variable the value of expression `f` depending on public variables; assignment to private variables or depending on the value of private variables is not allowed. The fourth statement assigns zero with probability x, and one with probability $1-x$, to a 1-bit public variable. The `return` statement outputs values of all public variables and terminates. A conditional branch first evaluates an expression `g` dependent on private and public variables, and it jumps to label l_a if `g` is true and to label l_b otherwise. Since only a single variable scope exists, loops can be added in a standard way as syntactic sugar.

As a contribution, we present a method to compute information leakage of a program by analyzing the execution traces of the program. We introduce the

Markovian semantics of our language by means of a function computing the successors of each state. Then we explain how we perform a depth-first exploration of the traces of the system, obtaining a set Q of final states that represent all possible output states of the system. Finally, we show how to compute the posterior entropy from Q.

$$\frac{pc\text{: } \texttt{public int}n\ \texttt{v}\ \texttt{:=}\ k}{succ(pc, L, H, p) = \{(pc+1, L \cup \{L(\texttt{v}) = k\}, H, p)\}}$$

$$\frac{pc\text{: } \texttt{private int}n\ \texttt{v}}{succ(pc, L, H, p) = \{(pc+1, L, H \cup \{H(\texttt{v}) = \{0, ..., 2^n - 1\}\}, p)\}}$$

$$\frac{pc\text{: } \texttt{skip}}{succ(pc, L, H, p) = \{(pc+1, L, H, p)\}} \qquad \frac{pc\text{: } \texttt{v := f}(L)}{succ(pc, L, H, p) = \{(pc+1, L \cup \{L(\texttt{v}) = \texttt{f}(L)\}, H, p)\}}$$

$$\frac{pc\text{: } \texttt{v := rand}\ x}{succ(pc, L, H, p) = \{(pc+1, L \cup \{L(\texttt{v}) = 0\}, H, p \cdot x), (pc+1, L \cup \{L(\texttt{v}) = 1\}, H, p \cdot (1-x))\}}$$

$$\frac{pc\text{: } \texttt{goto}\ label}{succ(pc, L, H, p) = \{(label, L, H, p)\}} \qquad \frac{pc\text{: } \texttt{return}}{succ(pc, L, H, p) = \emptyset}$$

$$\frac{pc\text{: } \texttt{if g}(L, H)\ \texttt{then goto}\ l_a\ \texttt{else goto}\ l_b}{\begin{array}{c} succ(pc, L, H, p) = \{(l_a, L, H|\texttt{g}(L, H), p \cdot Pr(\texttt{g}(L, H)|\pi(h))), \\ (l_b, L, H|\neg\texttt{g}(L, H), p \cdot Pr(\neg\texttt{g}(L, H)|\pi(h)))\} \end{array}}$$

Fig. 1. Successor function for each state in the Markovian trace semantics.

Semantics. The Markovity of the semantics allows us to define states containing enough information to determine a probability distribution over all traces originating from any state.

Definition 1. *A* state *in a Markovian semantics is a tuple* (pc, L, H, p) *where* $pc \in \mathbb{N}^0$ *is the program counter,* L *a set of assignments of values to public variables,* H *an set of assignments of sets of values to private variables, and* $0 \leq p \leq 1$ *is the probability of the state.*

The *initial state* of the semantics is $(1, \emptyset, \emptyset, 1)$. The *set of successor states* of a state (pc, L, H, p) depends on the statement pointed at by the program counter pc. States pointing to a `return` statement have 0 successors, states pointing to a `rand` or `if` statement have up to 2 successors, and any other state has 1 successor. The successor function defining the semantics of the language is shown in Fig. 1. If a state has zero probability, e.g. when a conditional is always true, it is removed from the set of successors.

We call a state *final* if it has no successors, meaning that the program counter of the state points to a `return` statement. The trace analysis terminates when a final state is encountered. This means that the analysis terminates if and only if the program under analysis terminates, so non-terminating programs cannot

be analyzed with this technique. Non-termination of the program under analysis raises other issues in leakage computation [4], and is not considered here.

Conditional states and random assignment states have two successors. The successors of a conditional state correspond to the guard being true or false. Since the guard can depend on the secret, both successor states may have positive probability depending on the prior distribution $\pi(h)$ on the secret, which is available at this time. The successors of a random assignment state correspond to the bit being set to 0 or 1. In both cases the probability of each successor state is computed and one of the successor states with non-zero probability is chosen to be the next step in the analysis. Successors with probability zero are dropped, pruning unreachable leaves from the trace tree.

Because of the Markovian semantics, each state contains the information to compute the probability distribution over its outgoing transitions. The probability of a trace is computed as the product of the probabilities of the transitions composing the trace. In the successor states of the conditional statement, $H|_{\mathtt{g}}(L,H)$ (resp. $H|_{\neg\mathtt{g}}(L,H)$) represents the assignment function obtained by removing from the sets of values assigned to the private variables those values that contradict (resp. respect) the guard $\mathtt{g}(L,H)$. Similarly, $Pr(\mathtt{g}(L,H)\,|\pi(h))$ (resp. $Pr(\neg\mathtt{g}(L,H)\,|\pi(h))$) refers to the probability that the guard $\mathtt{g}(L,H)$ is true (resp. false) considering the prior probability distribution $\pi(h)$ on the private variables.

When the analysis of a single trace terminates, the corresponding final state $(\bar{pc}, \bar{L}, \bar{H}, \bar{p})$ is produced, in which pc points to a `return` statement. The sets of allowed values assigned to the private variables in $\bar{H}$ have been appropriately reduced to account for the conditional statements visited by the trace.

Depth-First Trace Exploration. We perform a depth-first exhaustive exploration of the execution traces of the system, starting from the initial state $(1, \emptyset, \emptyset, 1)$. Each trace is explored until it gets to a final state, then the final state gets added to the multiset Q of final states. When all traces have been explored, the full multiset of final states Q of the system is produced. We then use Q to compute the posterior entropy of the system using Algorithm 1 presented below. The leakage of the system is computed as the difference between the prior and posterior entropy, as explained in Sect. 2.

Note that the exploration also depends on the prior distribution $\pi(h)$: values of the secret with a probability zero in the prior distribution are not explored. This behavior is intended, as is avoids unnecessarily exploring traces that have probability zero.

The depth-first exploration algorithm can be parallelized to take advantage of multicore architectures and is implemented in the current release of QUAIL, available at http://project.inria.fr/quail. Since this new algorithm is hundreds of times faster than the previous QUAIL implementation, we consider is as the standard QUAIL algorithm.

Posterior Uncertainty Computation. We show how to compute the posterior uncertainty $U(\pi(h|o))$ of a system with a secret `h` and an observable `o`,

Data: uncertainty measure U, multiset Q of final states
Result: posterior uncertainty $U(\pi(h|o))$
1 Initialize $\pi(o)$ and all $\pi(h, o = \bar{o})$ to zero;
2 **forall the** $s = (pc, L, H, p) \in Q$ **do**
3 Let $\bar{o} = L(o)$, $\{k_1, ..., k_n\} = H(h)$;
4 Set $\pi(o = \bar{o}) \leftarrow \pi(o = \bar{o}) + p$;
5 **for** $i = 1...n$ **do**
6 Set $\pi(h = k_i, o = \bar{o}) \leftarrow \pi(h = k_i, o = \bar{o}) + p/n$;
7 **end**
8 **end**
9 For each $\bar{o} \in D(o)$ let $\pi(h|o = \bar{o}) \leftarrow \pi(h, o = \bar{o})/\pi(o = \bar{o})$;
10 Return $U(\pi(h|o)) = \sum_{\bar{o} \in D(o)} \pi(o = \bar{o}) U(\pi(h|o = \bar{o}))$

Algorithm 1. Posterior uncertainty computation

given the uncertainty measure U and a multiset Q of final states of the system. Q encodes the posterior joint probability of all variables in the system and can be produced by the depth-first exploration algorithm presented above.

Let (pc, L, H, p) be a final state in Q, where L represents the assignments of given values to the public variables, H the assignments of sets of values to the private variables, and p the joint probability of such assignments. Since different traces may produce the same final assignments to variables (L, H), the joint probability of these assignments is the sum of the probabilities of all such final states. To apply the formula (2) $U(\pi(h|o)) = \sum_{\bar{o} \in D(o)} \pi(o = \bar{o}) U(\pi(h|o = \bar{o}))$, we need to compute the marginal probability distribution $\pi(o)$ and for each observable output $\bar{o} \in D(o)$ s.t. $\pi(o = \bar{o}) > 0$ the corresponding conditional probability distribution on h, i.e. $\pi(h|o = \bar{o})$.

Algorithm 1 computes $\pi(o)$ and each $\pi(h|o = \bar{o})$ by analyzing a multiset of final states. For each state (pc, L, H, p) the value of the observable variable $\mathtt{o}$ and set of values of the secret variable $\mathtt{h}$ are analyzed (lines 2–8). The probability of observing the value $\bar{o}$ of the observable variable in the state is increased by p (line 4), and the probability of observing each of the n values of the secret variable conditioned on $\bar{o}$ is increased by p/n (line 6). Finally, the probability on each subdistribution $\pi(h, o = \bar{o})$ is normalized to 1 by dividing it by $\pi(o = \bar{o})$ to obtain the conditional probability $\pi(h|o = \bar{o})$ (line 9) since $P(X|Y) = {}^{P(X,Y)}/_{P(Y)}$.

Theorem 1. *Algorithm 1 terminates and outputs the posterior uncertainty* $U(\pi(h|o))$ *of the posterior distribution represented by* Q.

3.2 LeakWatch

LeakWatch [13] estimates the leakage of a Java program with secrets and observations by running it several times for each possible value of the secret and inferring a probability distribution on the observations for each secret. The tool automatically terminates the analysis when the precision of the estimation is deemed sufficient, but different termination conditions can be used.

For small secrets, LeakWatch gives reliably approximates the leakage of complex Java programs. For larger secret, i.e. more than 10 bits, LeakWatch takes more time to return a value. However, the user can decide an acceptable error level for the tool to reduce the computation time necessary to obtain an answer. Also, if the tool is terminated prematurely, it can still provide an answer, even if it will be potentially quite imprecise. This makes LeakWatch the only tool of the three considered that can always provide an answer in a time-limited scenario, since QUAIL and Moped-QLeak generate a leakage result only if they complete their execution.

Finally, LeakWatch provides many command-line options for tuning the analysis parameters. In particular, one of the options displays the current estimation of the leakage at regular intervals, which can be very useful when developing.

Syntax and Usage. The syntax is the same as the Java language, with the additional commands `secret(name,value)` to declare a secret with a given name and value, and `observe(value)` to declare an observation of a given value. The analysis evaluates how much information leaks from the secret to the observable values. In particular, LeakWatch can compute leakage from a point of a program to another point of the program, and not necessarily from the start to the termination of the program.

To run LeakWatch, a Java programa annotated with `secret` and `observable` statements has to be compiled linking the LeakWatch library:

```
javac -cp leakwatch-0.5.jar:. MyClass.java
```

The tool is then run passing the name of the compiled class as a parameter:

```
java -jar leakwatch-0.5.jar MyClass
```

The tool returns its leakage estimate for the Java program. Normally LeakWatch determines automatically when it has run enough executions. We have used the `-n` parameter to fix the number of executions of the program when we experimented with different precisions and computation times.

3.3 Moped-QLeak

Moped-QLeak [11] uses the Moped tool [17] to compute a symbolic Algebraic Decision Diagram (ADD) representation of the *summary* of a program, which contains the relation between the inputs and outputs of the program. Moped-QLeak then computes Shannon or min-entropy leakage from this ADD representation using two algorithms introduced by the authors. To obtain the ADD representation of the program, Moped basically performs a fix-point iteration.

Moped's ability to build a symbolic representation of a program depends on the program's complexity. When such representation is computed, Moped-QLeak computes the information leakage with a small time overhead. On the other hand, some programs are not easy to reduce to a symbolic representation, and in this case Moped-QLeak's computation does not terminate within a reasonable time.

The ADD-based representation of probability distributions allows Moped-QLeak to analyze examples with large secret and observation spaces. In particular, the authors test it with 32-bit secrets and observables, whereas QUAIL's

computation time tends to be exponential in the size of the observables and LeakWatch's in the size of the secret. This suggests that the ADD approach is a key improvement on the state of the art, allowing the analysis tools to analyze off-the-shelf programs using 32- and 64-bit variables.

3.4 Syntax and Usage

The tool analyzes programs written in a variant of Moped's Remopla language. We provide here a simplified version of the syntax used by Moped-QLeak.

stmt ::= `skip ;` | *ident* `=` *exp*`;` | `pchoice (::`*prob*`->`*stmt*`)`$^+$ `choicep`
| `do ::` *exp* `->` *stmt* `::` `else ->` *stmt* `od`
| `if ::` *exp* `->` *stmt* `::` `else ->` *stmt* `fi`

The `if` and `do` constructs from Remopla, originally non-deterministic in Moped, have been made deterministic in Moped-QLeak. The language has also been enriched with a probabilistic choice operator, `pchoice` which allows the programmer to probablistically define the next statement (e.g. by giving a probability *prob* to each statement). Remopla supports loops, arrays and integers of arbitrary size. The language is normally used to encode systems for model checking against temporal logics.

The language does not provide constructs to declare secrets and observables, but assume that all global variables are at the same time secret and observable. More precisely, the initial values are considered as the input and the final values as the output. In practice, a variable is made secret by assigning it the same value in all final states.

Moped-QLeak is executed on a Remopla file `MyFile.rem` by calling

```
mql -shannon MyFile.rem
```

where `-shannon` specifies that the tool will compute and return the Shannon leakage.

4 Case Studies

We evaluate the three tools described in the previous Section with two scalable case studies[2]. The case studies have been chosen because they model real-life systems and the results computed are representative of realistic security concerns. In order to compare them, we consider the following criteria:

Speed. Evaluating the time required by the tool to provide a result;
Accuracy. Evaluating the precision of the result returned by the tool;
Scalability. Evaluating how the tool behaves on larger instances of the case studies;
Usability. Evaluating the easiness of modeling and the usefulness of the error messages from the compiler.

[2] The files used for our experiments are available at https://project.inria.fr/quail/casestudies/.

4.1 Case Study A: Smart Grids

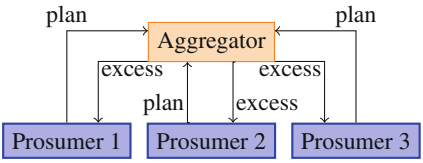

Fig. 2. Smart grid overview

A Smart grid is an energy network where every node may produce, store and consume energy. Nodes are called *prosumers* (PROducer consSUMERS). The *Living Lab* demonstrator [22] is an instance of such a prosumer, whose data can be accessed online[3]. The prosumers periodically negotiate with an aggregator in charge of balancing the consumption and production among several prosumers. Figure 2 depicts a grid with 3 prosumers. Each prosumer declares its plan, that is, how much it intends to consume or produce during the next period of time. The aggregator sends to each house the value indicating the excess of energy production or consumption. An excess of 0 indicates that the plans are feasible and terminates the negotiation. Otherwise, the prosumers adapt their plan accordingly and send the updated version. Smart grid and smart sensors raise several security and privacy concerns. The platform can ensure the information cannot flow directly between prosumers [10]. However, stability requires a feedback from the aggregator that potentially conveys information about other prosumer, where only the software can limit information leakage. In general, knowing the consumption of a particular household may reveal some sensitive information about the house (presence of people in the house, type of electrical devices ...). Therefore, the consumption of a prosumer should remain secret. The privacy of a prosumer with respect to the aggregator can be ensured in several ways [29]. However, each prosumer receives some information about the consumption of other prosumers through the excess value sent back by the aggregator.

An attacker might use the information obtained through the grid in order to decide whether a given house is occupied or not. In our scenario, we assume different types of houses with different consumptions. Each house is modeled by a private boolean value, which is true if the house is occupied. An occupied house consumes a fixed amount of energy, according to its type. An empty house does not consume anything. Table 1 presents how much a given house consumes, in two different cases that we consider.

Table 1. Consumption of houses wrt size

Size	Case A	Case B
Small	1	1
Medium	2	3
Large	3	5

For this experiment, we assume that the attacker observes the global consumption of the quarter. We consider different targets for the attack and thus different secrets. Either the attacker targets a single house of a given type (i.e. S, M or L) and only the bit corresponding to the presence in that house is secret, or the attacker wants to obtain informations about all the houses and the whole array of bits indicating the presence in each house is secret.

[3] livinglab.fortiss.org.

Table 2. Leakage of presence information through the global consumption

Case	Nb of houses	Single house leakage			Global leakage	Global leakage/bit
		S	M	L		
A	3	0.7500	0.7500	0.7500	2.7500	0.9166
A	6	0.0688	0.1466	0.2944	3.4210	0.5701
A	9	0.0214	0.0768	0.1771	3.7363	0.4151
A	12	0.0135	0.0544	0.1273	3.9479	0.3289
B	3	1.0000	1.0000	1.0000	3.0000	1.0000
B	6	0.1965	0.1965	0.3687	4.0243	0.6707
B	9	0.0241	0.0808	0.2062	4.3863	0.4873
B	12	0.0074	0.0510	0.1443	4.6064	0.3838

Table 3. Time to compute or approximate the leakage for a large house

Case	House Nb	Time QUAIL	Time LW	Time mql
A	3	0.1 s	0.3 s	0.02 s
A	6	0.3 s	0.3 s	0.02 s
A	9	0.6 s	0.4 s	0.02 s
A	12	1.6 s	0.4 s	0.03 s
B	3	0.2 s	0.3 s	0.02 s
B	6	0.3 s	0.5 s	0.02 s
B	9	0.6 s	0.4 s	0.02 s
B	12	1.7 s	0.4 s	0.03 s

Usability. We model the above scenario in the three tools. We consider two versions depending on the target. The program is rather simple to model, the only noticeable difference between the tools language is the declaration of unobservable variables. When targeting all the houses, the secret is an array of boolean. When targeting a single house, the secret is a single boolean. Both targets are supported by all the tools. However, the presence in the other houses is not a secret, but still an unknown and unobservable input of the program. In QUAIL, the `private` keyword allows the programmer to declare directly such variables. With LeakWatch, we chose these values randomly but do not declare them as secret. In Moped-QLeak, we choose these values randomly, as in LeakWatch.

Table 2 presents the leakage for the Smart grid case study. The first two columns indicate the case, as presented in Table 1 and the number of houses in the model. For a model with N houses, there are $N/3$ houses of each type. The columns S, M and L indicates the leakage of the variable representing the presence in a house of the corresponding type. The column "Global leakage" contains the leakage of the whole array of presence information bits and the column "Global leakage/bit" indicates the average leak per bit of secret.

In Case B with only 3 houses, the presence information can be deduced from the global consumption information, which is indicated by a leakage of 1 for each presence bit. Otherwise, the average leakage per bit from a global attack

Table 4. Average relative error and computation time over 100 runs for computing the leakage of the presence in a large house within 12 houses in Case B.

Tool	`mql`	QUAIL	LeakWatch						
Nb. of Simulations	-	-	Default	1000	2000	5000	10000	20000	50000
Error	0 %	0 %	14.0 %	10.4 %	6.4 %	4.8 %	2.8 %	2.1 %	1.4 %
Time	0.031 s	1.7 s	0.4 s	0.7 s	1.0 s	2.1 s	3.7 s	6.9 s	16.6 s

is more important that the information obtained by focusing on a single house. This means than obtaining information about the whole array, for instance the number of occupied houses, is easier than obtaining information about a single bit, i.e. presence information of a single house. In both cases, the leakage, and thus the loss of anonymity of prosumers, diminishes when the number of houses increases.

Speed. In Table 3 we show the time needed by QUAIL, LeakWatch and Moped-QLeak for computing the leakage of the presence information in a house of size L. Moped-QLeak takes around 20 ms to compute this value, LeakWatch takes between 300 and 500 ms and QUAIL takes between 100 and 1700 ms, depending on the size of the model. Furthermore, Moped-QLeak and QUAIL compute the exact leakage value, whereas LeakWatch computes an approximation. For a more precise comparison, we need to take precision into account.

Accuracy. We compare QUAIL, LeakWatch and Moped-QLeak on computing the leakage of the presence information of a single large house, in Case B. QUAIL takes 1.7 s to compute the exact leakage. With the default parameters, LeakWatch takes 0.4 s to compute an approximation with a relative error of 14 % (average on 100 runs). It requires 500 to 700 simulations.

To compare execution times with respect to errors, we did an additional experiment, where we requested LeakWatch to run more simulations. For each requested number of simulations we provide in Table 4 the average relative error (over 100 runs) and the time needed for the computation. We see that for an equivalent amount of time, LeakWatch provides a result with a relative error of 4 to 6 %, whereas QUAIL returns the exact result. Moped-QLeak is the fastest and most precise.

Scalability. Finally, we evaluate the scalability of the tools by increasing the number of houses until the analysis time reaches 1 h. For this experiment, we evaluate the leakage of the presence information, in Case B, for a single house of size L (1 bit secret), or for all the houses simultaneously (N bits of secret). The results are shown in Table 5.

We see that LeakWatch can handle a very large number of houses when computing the leakage from a small secret, but is not much more scalable than QUAIL with a large secret. Recall that LeakWatch provide an approximation of the leakage, whereas QUAIL and Moped-QLeak provide the exact value. Moped-QLeak scales relatively well with both a small and a large secret to analyze.

Table 5. Maximal size analyzable in one hour

Target	LW	QUAIL	mql
L-size house	150000	27	234
All houses	15	12	150

4.2 Case Study B: Voting Protocols

In an election, each voter is called to express his preference for the competing candidates. The *voting system* defines the way the voters express their preference: either on paper in a traditional election, or electronically in e-voting. After

the votes have been cast, the *results* of the vote are published, usually in an aggregated form to protect the anonymity of the voters. Finally, the winning candidate or candidates is chosen according to a given *electoral formula*.

In this section we present two typologies of voting, representing two ways in which the voters can express their preference: in the *Single Preference* protocol the voters declare their preference for exactly one of the candidates, while in the *Preference Ranking* protocol each voter ranks the candidate from his most favorite to his least favorite.

Single Preference. This protocol typology models all electoral formulae in which each of the N voters expresses one vote for one of the C candidates, including plurality and majority voting systems and single non-transferable vote [24]. The votes for each candidate are summed up and only the results are published, thus hiding information about which voter voted for which candidate. The candidate or candidates to be elected are decided according to the electoral formula used.

The secret is an array of integers with a value for each of the N voters. Each value is a number from 0 to $C-1$, representing a vote for one of the C candidates. The observable is an array of integers with the votes obtained by each of the C candidates.

The protocol is simple, and its information leakage can be computed formally, as shown by the following lemma:

Lemma 1. *The information leakage for the Single Preference protocol with n voters and c candidates corresponds to*

$$-\sum_{k_1+k_2+\ldots+k_c=n} \frac{n!}{c^n k_1!k_2!\ldots k_c!} \log_2\left(\frac{n!}{c^n k_1!k_2!\ldots k_c!}\right)$$

While the lemma provides a formula to "manually" compute the leakage, it is very hard to find such a formula for an arbitrary process. Therefore automated tools should be employed.

Preference Ranking. This protocol typology models all electoral formulae in which each of the n voters expresses an order of preference of the c candidates, including the alternative vote and single transferable vote systems [24]. In the Preferential Voting protocol the voter does not express a single vote, but rather a ranking of the candidates; thus if the candidates are A, B, C and D the voter could express the fact that he prefers B, then D, then C and finally A. Then each candidate gets c points for each time he appears as first choice, $c-1$ points for each time he appears as second choice, and so on. The points of each candidate are summed up and the results are published.

The secret is an array of integers with a value for each of the N voters. Each value is a number from 0 to $C!-1$, representing one of the possible $C!$ rankings of the C candidates. The observable is an array of integers with the points obtained by each of the C candidates.

Table 6. Voting protocols: percent of secret leaked by Single Preference (on the left) and Preference Ranking (on the right) computed with the QUAIL tool. Timeout is set at 1 h.

SP		Candidates				
		2	**3**	**4**	**5**	**6**
Voters	**3**	60.4 %	65.7 %	69.0 %	71.3 %	73.0 %
	4	50.8 %	56.5 %	60.2 %	62.9 %	64.9 %
	5	44.0 %	49.6 %	53.5 %	56.4 %	58.6 %
	6	38.9 %	44.4 %	48.3 %	51.2 %	53.5 %

PR		Candidates		
		2	**3**	**4**
Voters	**3**	60.4 %	61.9 %	62.0 %
	4	50.8 %	51.0 %	timeout
	5	44.0 %	43.4 %	timeout
	6	38.9 %	37.9 %	timeout

Experimental Results

Usability. We model the two voting systems, where the secret is the votes, and the observable the results. In single preference voting, the secret is an array of integer that represent individual votes. The range of this integer corresponds to the number of candidates. In QUAIL, it is possible to declare the range of a secret integer. In LeakWatch, each vote is drawn uniformly in the range and then declared secret. In Moped-QLeak, this case requires more work. The range of a secret integer depends on the chosen size bits. A special variable, `out_of_domain`, is set to true if one of the votes is not in the valid range and the corresponding input is not considered. Furthermore, when using this variable, it's not possible to use local variables, which is indicated by the error message "The first computed value is not a constant.". The impossibility to use local variables and the imprecision of the error message increased considerably the modelling time.

For the Preferential Voting, we were not able to produce a Moped-QLeak program that terminates. We suspect that Moped is unable to compute a symbolic representation of the Preferential Voting protocol due to its inherent complexity. Indeed, this program decodes an integer between 0 and the factorial of the number of candidates into a sorted list of the candidates, to assign the corresponding points to the candidates.

Accuracy. Table 6 shows the percentage of the secret leaked by the Single Preference and Preference Ranking protocols for different numbers of voters and candidates. The results for 2 candidates are identical, since in this case in both protocols the voters can vote in only 2 different ways. The results obtained for Single Preference are correct with respect to the formula stated in Lemma 1. The table shows that the Single Preference protocol leaks a larger part of its secret than the Preference Ranking protocol.

Table 7 shows the percent error of the leakage value obtained with LeakWatch. Indeed, LeakWatch computes an approximation of the leakage based on simulation, whereas QUAIL and Moped-QLeak compute the exact value. Furthermore, the leakage computed by LeakWatch for a given program may change

Table 7. Percent error of the leakage obtained by LeakWatch relatively to the exact value for Single Preference (on the left) and Preference Ranking (on the right). Timeout is set to 1 h.

SP	Candidates				
Voters	2	3	4	5	6
3	-3.8 %	-3.7 %	-3.2 %	-2.8 %	-2.2 %
4	-5.1 %	-3.7 %	-2.6 %	-2.3 %	-2.1 %
5	-5.0 %	-3.2 %	-2.6 %	-2.2 %	-1.9 %
6	-5.1 %	-3.2 %	-2.4 %	timeout	timeout

PR	Candidates		
Voters	2	3	4
3	-3.8 %	-2.6 %	timeout
4	-5.1 %	-2.6 %	timeout
5	-5.0 %	-2.2 %	timeout
6	-5.1 %	timeout	timeout

Table 8. Time in seconds needed to compute the leakage for Single Preference with QUAIL (left), LeakWatch (middle) and Moped-QLeak (right). Timeout is set to 1 h.

SP	Candidates				
QUAIL	2	3	4	5	6
Voters 3	0.2	0.3	0.4	0.5	0.8
Voters 4	0.3	0.5	1.0	1.6	2.7
Voters 5	0.3	0.9	2.5	6.8	13.3
Voters 6	0.5	2.8	13.3	56.7	214.4

SP	Candidates				
LW	2	3	4	5	6
Voters 3	0.4	0.8	2.5	6.9	19.1
Voters 4	0.5	2.4	14.1	64.6	232.3
Voters 5	0.7	8.1	81.6	549.4	2688.3
Voters 6	1.1	29.0	481.6	to	to

SP	Candidates				
mql	2	3	4	5	6
Voters 3	0.8	0.8	0.9	1.0	1.1
Voters 4	0.9	0.5	0.9	1.2	1.6
Voters 5	1.0	1.1	1.2	6.8	2.7
Voters 6	1.1	1.2	1.6	2.5	4.6

at each invocation of the tool, because LeakWatch samples random executions. Here, LeakWatch slightly underestimates the leakage, by 2 to 5 %.

Speed. We compare the execution time of the three tools in Table 8 for Single Preference and in Table 9 for Preference Ranking. These execution times have been obtained on a laptop with a i7 quad-core running at 3.3 GHz and 16 GB of RAM. The results show that QUAIL is significantly faster than LeakWatch on these examples. This shows that QUAIL performs better than LeakWatch with large secrets, in line with previous results [5]. For single preference, Moped-QLeak clearly outperforms QUAIL on large examples. The results for Moped-QLeak in the preferential voting case studies are missing from Table 9 because the tool did not terminate in this case study, even with the smallest instance of 2 voters and 2 candidates.

Scalability. Concerning the Scalability, we see that QUAIL and Moped-Qleak are more scalable than LeakWatch, since the latter times out in Tables 8 and 9. For Single Preference, QUAIL stops at 7 voters and 6 candidates, due to an error. Moped-QLeak finished with 12 voters and 6 candidates but returned `-inf` as leakage value, instead of 11. With 9 voters and 6 candidates, the result has approximately 1 bit of errors. Therefore, we conjecture that the `-inf` value is

Table 9. Time in seconds needed to compute the leakage for Preference Ranking with QUAIL (on the left) and LeakWatch (on the right). Timeout is set to one hour.

PR		Candidates		
QUAIL		**2**	**3**	**4**
Voters	**3**	0.3	2.0	89.4
	4	0.4	9.0	timeout
	5	0.7	76.7	timeout
	6	1.1	2987.8	timeout

PR		Candidates		
LW		**2**	**3**	**4**
Voters	**3**	0.4	13.7	timeout
	4	0.5	121.0	timeout
	5	0.8	1267.3	timeout
	6	1.2	timeout	timeout

a precision error. On these examples, no tool seems to be much more scalable than the others, due to various reasons.

5 Conclusions

In this paper, we provided two scalable case studies for the leakage computation and used them for comparing the existing tools able to perform such an approximation. We have compared the state of the art in information leakage tools – LeakWatch, QUAIL and Moped-QLeak – on their speed, accuracy, scalability and usability in addressing the case studies. We summarize here our observations and experience with the tools.

Speed. Concerning the execution time, Moped-QLeak is usually the fastest tool in providing an exact result. However, in the preferential voting example Moped-QLeak was unable to terminate its analysis in less than one hour even for the smallest instances of the problem. We can note that LeakWatch is faster than QUAIL on small secrets (e.g. 1 bit) but QUAIL outperforms LeakWatch on larger secrets. Finally, LeakWatch is very fast on small secrets, but its result and evaluation of the system (presence or absence of leakage) tends to change between different executions of the tool.

Accuracy. The tool giving the most accurate result is QUAIL because it supports arbitrary precision. LeakWatch provides an approximated result and therefore is imprecise by definition. Moped-QLeak does not implement arbitrary precision analysis, and consequently suffers from approximation errors. For instance, we found an error in the order of 1 bit on the majority voting protocol with 9 voters and 6 candidates, for which we have the exact result. Also, for the same protocol with 12 voters and 6 candidates Moped-QLeak reported a leakage of negative infinity bits, which we conjecture is caused by approximation and division-by-zero errors in the computation.

Scalability. For small secrets, LeakWatch scales better than the other tools analyzed. In the Smart Grid case study, we managed to analyze the leakage for an

aggregation of 150000 houses in less than one hour.However, the returned result is obtained statistically, and varies from one execution to the other.

For large secrets, the winner is Moped-QLeak, as it scales much better than QUAIL on the Smart Grid case study. However, for the voting protocol, QUAIL manages to analyze only two voters less than Moped-QLeak (6 against 8), before approximation issues make Moped-QLeak's results incorrect.

Usability. Since all the tools studied here are academic tools who are still in their early years, usability is not necessarily the main concern of their developers. However, we have found some important discrepancies in this area.

The most usable tool is LeakWatch, especially if the program to analyze is already written in Java. In that case, it is sufficient to annotate the program in order to declare the secrets and the observable values. Furthermore, LeakWatch has a command line option to display the current results based on the traces collected so far, which is convenient when the analysis time is very long.

QUAIL has its own language, which is an imperative WHILE language with arrays and constants. QUAIL allows the explicitly declaration of variables as observable, public, private or secret, with a specific range of allowed values. Furthermore, QUAIL has a command-line option to change the values of constants declared in a program, which comes in handy when performing batch experimentation.

Using Moped-QLeak has been more problematic because of some issues with the Remopla language. In particular, the range of the secrets cannot be determined, instead the program has to raise an `out_of_domain` exception when the values are not in the expected range. Also, all integer variables have the same length, defined in the `DEFAULT_INT_BITS` constant. Finally, some error messages are misleading and slow down the modelling process.

To conclude, Moped-QLeak is the fastest tool, because it uses a suitable data structure (Algebraic Decision Diagrams) for representing the executions. However, this data structure may become a problem with complex program, as shown by the preference ranking example, which Moped-QLeak cannot analyze, contrarily to the other tools. The other tools, QUAIL and LeakWatch are more usable. QUAIL, which also has its own dedicated language, provide some specific constructs for declaring the visibility and range of a variable.

We believe that reimplementing QUAIL with a better data structure for probability distributions, like the ADDs used in Moped-QLeak, would provide a fast and usable tool for performing leakage analysis. The statistical techniques used in LeakWatch should also be integrated to allow approximated results for large instances.

6 Related Tools

We discuss some security-related automated tools and their relation with the work presented in this paper.

The STA tool developed by Boreale et al. [7] is similar in intent to the algorithms we propose, since it also uses symbolic trace analysis. More recent

work by Boreale et al. [8] introduces a semiring-based semantics able to perform compositional quantitative analysis of non-deterministic systems, but no tool is available at the moment.

Efficient tools have been developed by Phan and Malacaria for information-theoretical analysis of systems. The tools squifc [25], QILURA [26], and jpf-qif [27] use SMT solving to perform a symbolic analysis of C or Java code and to compute channel capacity of programs, where the channel capacity is the maximum information leakage achievable for any prior distribution over the secret and randomness of the system. Since the tools compute channel capacity and not Shannon leakage of randomized systems, they have not been included in our comparison.

McCamant et al. have obtained interesting results in detecting leakage of information by implicit flow by applying dynamic and quantitative taint analysis techniques [20,23]. Again, their techniques have not been included in this evaluation since they do not compute information-theoretical leakage measures like Shannon and min-entropy leakage.

References

1. Alvim, M.S., Chatzikokolakis, K., Palamidessi, C., Smith, G.: Measuring information leakage using generalized gain functions. In: Chong, S. (ed.) CSF, pp. 265–279. IEEE (2012)
2. Backes, M., Doychev, G., Köpf, B.: Preventing side-channel leaks in web traffic: A formal approach. In: NDSS. The Internet Society (2013)
3. Biondi, F., Legay, A., Malacaria, P., Wąsowski, A.: Quantifying information leakage of randomized protocols. In: Giacobazzi, R., Berdine, J., Mastroeni, I. (eds.) VMCAI 2013. LNCS, vol. 7737, pp. 68–87. Springer, Heidelberg (2013)
4. Biondi, F., Legay, A., Nielsen, B.F., Malacaria, P., Wasowski, A.: Information leakage of non-terminating processes. In: Raman and Suresh [28], pp. 517–529
5. Biondi, F., Legay, A., Traonouez, L.-M., Wąsowski, A.: QUAIL: a quantitative security analyzer for imperative code. In: Sharygina, N., Veith, H. (eds.) CAV 2013. LNCS, vol. 8044, pp. 702–707. Springer, Heidelberg (2013)
6. Boreale, M.: Quantifying information leakage in process calculi. Inf. Comput. **207**(6), 699–725 (2009)
7. Boreale, M., Buscemi, M.G.: Experimenting with STA, a tool for automatic analysis of security protocols. In: SAC, pp. 281–285. ACM (2002)
8. Boreale, M., Clark, D., Gorla, D.: A semiring-based trace semantics for processes with applications to information leakage analysis. Math. Struct. Comput. Sci. **25**(2), 259–291 (2015)
9. Boreale, M., Pampaloni, F.: Quantitative information flow under generic leakage functions and adaptive adversaries. In: Ábrahám, E., Palamidessi, C. (eds.) FORTE 2014. LNCS, vol. 8461, pp. 166–181. Springer, Heidelberg (2014)
10. Bytschkow, D., Quilbeuf, J., Igna, G., Ruess, H.: Distributed MILS architectural approach for secure smart grids. In: Cuéllar [15], pp. 16–29
11. Chadha, R., Mathur, U., Schwoon, S.: Computing information flow using symbolic model-checking. In: Raman and Suresh [28], pp. 505–516
12. Chothia, T., Guha, A.: A statistical test for information leaks using continuous mutual information. In: CSF, pp. 177–190. IEEE Computer Society (2011)

13. Chothia, T., Kawamoto, Y., Novakovic, C.: LeakWatch: estimating information leakage from Java programs. In: Kutyłowski, M., Vaidya, J. (eds.) ICAIS 2014, Part II. LNCS, vol. 8713, pp. 219–236. Springer, Heidelberg (2014)
14. Chothia, T., Kawamoto, Y., Novakovic, C., Parker, D.: Probabilistic point-to-point information leakage. In: CSF, pp. 193–205. IEEE (2013)
15. Cuéllar, J. (ed.): SmartGridSec 2014. LNCS, vol. 8448. Springer, Heidelberg (2014)
16. Denning, D.E.: A lattice model of secure information flow. Commun. ACM **19**(5), 236–243 (1976)
17. Esparza, J., Kiefer, S., Schwoon, S.: Abstraction refinement with Craig interpolation and symbolic pushdown systems. J. Satisfiability, Boolean Model. Comput. **5**, 27–56 (2008). Special Issue on Constraints to Formal Verification
18. Goguen, J.A., Meseguer, J.: Security policies and security models. In: IEEE Symposium on Security and Privacy, pp. 11–20. IEEE Computer Society (1982)
19. Gray III, J.W.: Toward a mathematical foundation for information flow security. In: IEEE Symposium on Security and Privacy, pp. 21–35 (1991)
20. Kang, M.G., McCamant, S., Poosankam, P., Song, D.: DTA++: dynamic taint analysis with targeted control-flow propagation. In: NDSS. The Internet Society (2011)
21. Köpf, B., Mauborgne, L., Ochoa, M.: Automatic quantification of cache side-channels. In: Madhusudan, P., Seshia, S.A. (eds.) CAV 2012. LNCS, vol. 7358, pp. 564–580. Springer, Heidelberg (2012)
22. Koss, D., Sellmayr, F., Bauereiß, S., Bytschkow, D., Gupta, P.K., Schätz, B.: Establishing a smart grid node architecture and demonstrator in an office environment using the SOA approach. In: SE4SG. ICSE, pp. 8–14. IEEE (2012)
23. Newsome, J., McCamant, S., Song, D.: Measuring channel capacity to distinguish undue influence. In: Chong, S., Naumann, D.A. (eds.) PLAS. ACM (2009)
24. Norris, P.: Electoral Engineering: Voting Rules and Political Behavior. Cambridge University Press, Cambridge (2004). Cambridge Studies in Comparative Politics
25. Phan, Q., Malacaria, P.: Abstract model counting: a novel approach for quantification of information leaks. In: Moriai, S., Jaeger, T., Sakurai, K. (eds.) ASIACCS, pp. 283–292. ACM (2014)
26. Phan, Q., Malacaria, P., Pasareanu, C.S., d'Amorim, M.: Quantifying information leaks using reliability analysis. In: Rungta, N., Tkachuk, O. (eds.) SPIN, pp. 105–108. ACM (2014)
27. Phan, Q., Malacaria, P., Tkachuk, O., Pasareanu, C.S.: Symbolic quantitative information flow. ACM SIGSOFT Softw. Eng. Notes **37**(6), 1–5 (2012)
28. Raman, V., Suresh, S.P. (eds.) FSTTCS, vol. 29. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik (2014)
29. Rottondi, C., Fontana, S., Verticale, G.: A privacy-friendly framework for vehicle-to-grid interactions. In: Cuéllar [15], pp. 125–138
30. Smith, G.: On the foundations of quantitative information flow. In: de Alfaro, L. (ed.) FOSSACS 2009. LNCS, vol. 5504, pp. 288–302. Springer, Heidelberg (2009)