On the Automatic Construction of Indistinguishable Operations

M. Barbosa[1]* and D. Page[2]

[1] Departamento de Informática, Universidade do Minho,
Campus de Gualtar, 4710-057 Braga, Portugal.
mbb@di.uminho.pt

[2] Department of Computer Science, University of Bristol,
Merchant Venturers Building, Woodland Road,
Bristol, BS8 1UB, United Kingdom.
page@cs.bris.ac.uk

Abstract. An increasingly important design constraint for software running on ubiquitous computing devices is security, particularly against physical methods such as side-channel attack. One well studied methodology for defending against such attacks is the concept of indistinguishable functions which leak no information about program control flow since all execution paths are computationally identical. However, constructing such functions by hand becomes laborious and error prone as their complexity increases. We investigate techniques for automating this process and find that effective solutions can be constructed with only minor amounts of computational effort.

Keywords. Side-channel Cryptanalysis, Simple Power Analysis, Countermeasures, Indistinguishable Operations.

1 Introduction

As computing devices become increasingly ubiquitous, the task of writing software for them has presented programmers with a number of problems. Firstly, devices like smart-cards are highly constrained in both their computational and storage capacity; due to their low unit cost and small size, such devices are significantly less powerful than PDA or desktop class computers. This demands selection and implementation of algorithms which are sensitive to the demands of the platform. Coupled with these issues of efficiency, which are also prevalent in normal software development, constrained devices present new problems for the programmer. For example, one typically needs to consider the power characteristics and communication frequency of any operation since both eat into the valuable battery life of the device.

Perhaps the most challenging aspect of writing software for ubiquitous computers is the issue of security. Performing computation in a hostile, adversarial

* Funded by scholarship SFRH/BPD/20528/2004, awarded by the Fundação para a Ciência e Tecnologia, Ministério da Ciência e do Ensino Superior, Portugal.

environment demands that software is robust enough to repel attackers who hope to retrieve data stored on the device. Although cryptography provides a number of tools to aid in protecting the data, the advent of physical attacks such as side-channel analysis and fault injection mean one needs to consider security of the software implementation as well as the mathematics it implements. By passive monitoring of execution features such as timing variations [15], power consumption [16] or electromagnetic emission [1, 2] attackers can remotely recover secret information from a device with little fear of detection. Typically attacks consist of a collection phase which provides the attacker with profiles of execution, and an analysis phase which recovers the secret information from the profiles. Considering power consumption as the collection medium from here on, attack methods can be split into two main classes. Simple power analysis (SPA) is where the attacker is given only one profile and is required to recover the secret information by focusing mainly on the operation being executed. In contrast, differential power analysis (DPA) uses statistical methods to form a correlation between a number of profiles and the secret information by focusing mainly on the data items being processed.

As attack methods have become better understood, so have the related defence methods. Although new vulnerabilities are regularly uncovered, one can now deploy techniques in hardware and software which will vastly reduce the effectiveness of most side-channel attacks and do so with fairly minor overhead. Very roughly, defence methods fall into one of two main categories:

Randomisation One method of reducing the chance of leaking secret information is to introduce a confusion or randomisation element into the algorithm being executed. This is particularly effective in defending against DPA-style attacks but may also be useful in the SPA-style case. Essentially, randomisation ensures the execution sequence and intermediate results are different for every invocation and hence reduces the correlation of a given profile with the secret information. This method exists in many different forms, for example the addition of blinding factors to exponents; dynamically randomising the parameters or control flow in exponentiation algorithms; and using redundant representations.

Indistinguishability To prevent leakage of secret information to an SPA-style attack by revealing the algorithm control flow, this approach aims to modify operations sequences so that every execution path is uniform. Again, there are several ways in which this can be achieved. One way is to work directly on the mathematical formulae that define the operations and modify them so that the resulting implementations have identical structure. Another method is to work directly on the code, rearranging it and inserting dummy operations, to obtain the same effect.

A key difference between issues of efficiency and security is that the programmer is assisted by a compiler in the former case but not in the later. That is, the programmer is entirely responsible for constructing defence methods against side-channel analysis. Although the general technique of creating indistinguishable functions to foil SPA style attack is well understood; the general barrier

Algorithm 1 The double-and-add method for ECC point multiplication.

```
Input: point $P$, integer $d$
Output: point $Q = d \cdot P$
Q ← $\mathcal{O}$
for $i = |d| - 1$ downto 0 do
    $Q \leftarrow 2 \cdot Q$
    if $d_i = 1$ then
        $Q \leftarrow Q + P$
    end if
end for
return $Q$
```

to implementation is how labour intensive and error prone the process is. This is especially true when operation sequences in the functions are more complex than in the stock example of elliptic curve cryptography (ECC), for example systems like XTR or hyperelliptic curve cryptography (HECC). However, the task is ideally suited to automation; to this end our focus in this paper is the realisation of such automation to assist the development of secure software. In the rest of this Section we introduce the concept and use of indistinguishable functions in more detail and present an overview of related work. Then, in Section 2 we describe the construction of such functions as an optimisation problem and offer an algorithm to produce solutions in Section 3. Finally, we present some example results in Section 4 and concluding remarks in Section 5.

1.1 Using Indistinguishable Functions

One of the most basic forms of side-channel attack is that of simple power analysis (SPA): the attacker is presented with a single profile from the collection phase and tasked with recovering the secret information. Such an attack can succeed if one can reconstruct the program control flow by observing the operations performed in an algorithm. If decisions in the control flow are based on secret information, it is leaked to the attacker. We focus here on point multiplication as used in ECC [4] and described by Algorithm 1.

Restricting ourselves to working over the field $K = \mathbb{F}_p$, where p is a large prime, our elliptic curve is defined by:

$$E(K) : y^2 = x^3 + Ax + B$$

for some parameters A and B. The set of rational points $P = (x, y)$ on this curve, together with the identity element $\mathcal{O}$, form an additive group. ECC based public key cryptography typically derives security by presenting an intractable discrete logarithm problem over this curve group. That is, one constructs a secret integer d and performs the operation $Q = d \cdot P$ for some public point P. Since reversing this operation is believed to be hard, one can then transmit Q without revealing the value of d.

Point addition and doubling on an elliptic curve and often distinguishable from each other as one is composed from a different sequence of operations than the other. Denoting addition by A and doubling by D, the collection phase of a power based side-channel attack presents the attacker with a profile detailing the operations performed during execution of the algorithm. For example, by monitoring execution of using the multiplier $d = 1001_2 = 9_{10}$, one obtains the profile:

$$DADDDA$$

Given this single profile, the analysis phase can recover the secret value of d simply by spotting where the point additions occur. If the sequence DA occurs we have that $d_i = 1$ whereas if the sequence D occurs then $d_i = 0$.

One way to avoid this problem is to employ a double-and-add-always method, due to Coron [8], whereby a dummy addition is executed if the real one is not. Although the cases where $d_i = 0$ and $d_i = 1$ are now indistinguishable, this method significantly reduces the performance of the algorithm since many more additions are performed.

However, the ECC group law is very flexible in terms of how the point addition and doubling operations can be implemented through different curve parameterisations, point representations and so on. We can utilise this flexibility to force indistinguishability by manipulating the functions for point addition and doubling so that they are no longer different. This is generally achieved by splitting the more expensive point addition into two parts, each of which is identical in terms of the operations it performs to a point doubling. Put more simply, instead of recovering the profile above from the SPA collection phase, an attacker gets:

$$DDDDDDDD$$

from which they can get no useful information. Note that although we present the use of indistinguishable functions solely for point multiplication or exponentiation, the technique is more generally useful and can be applied in many other contexts.

1.2 Related Work

Gebotys and Gebotys [12] analyse the SPA resistance of a DSP-based implementation of ECC point multiplication using projective coordinates on curves over $\mathbb{F}_p$. They show that by hand-modifying the doubling and adding implementation code, simply by inserting dummy operations, it is possible to obtain significant improvements. Likewise, Trichina and Bellezza [21] analyse the overhead associated with the same approach using mixed coordinates on curves over $\mathbb{F}_{2^n}$, and again find an efficient hand-constructed solution. Brier and Joye [6] present unified addition and doubling functions by observing that operations for calculating slope can be shared between the two cases. Joye and Quisquater [14] and Liardet and Smart [18] take a different approach by finding different curve parameterisations that offer naturally indistinguishable formula; they utilise Hessian and Jacobi form elliptic curves respectively. In other contexts than ECC, Page and Stam [20] present hand-constructed indistinguishable operations for XTR.

Chevallier-Mames et al. [7] propose a generalised formulation for constructing indistinguishable functions and apply it to processor-level sequences of instructions. SPA attacks typically exploit conditional instructions that depend on secret information: the solution is to make the sequences of instructions (processes) associated with both branches indistinguishable. The authors introduce the concept of side channel atomicity: all processes are transformed, simply by padding them with dummy operations, so that they execute as a repetition of a small instruction sequence (a pattern) called a side-channel atomic block. This idea is closely related to our work and in some ways more powerful: one can hope to get better results from side-channel atomicity since it allows the instructions from one function to be mixed with those from another. This offers the potential for short patterns and hence efficient solutions to indestinguishability.

2 Indistinguishable Functions

In this section we enunciate the problem of building indistinguishable functions as an optimisation problem. We begin by defining a problem instance.

Definition 1. *Let F be a list of N functions $F = F_1, F_2, ..., F_N$ where each function F_i is itself a list of instructions from a finite instruction set L:*

$$F_i = F_i[1], F_i[2], ..., F_i[|F_i|]$$

where $|F_i|$ denotes the length of function F_i, and $F_i[j] \in L$ denotes instruction j of function F_i, with $1 \leq j \leq |F_i|$. Also, let $F_i[k..j]$ denote instructions k to j in function F_i, with $1 \leq k \leq j \leq |F_i|$.

For concreteness one should think of the simple case of two functions F_1 and F_2 as performing ECC point addition and doubling. Further, the instruction set L is formed from three-address style operations [19] on elements in the base field, for example addition and multiplication, and the functions are straight-line in that they contain no internal control flow structure.

We aim to manipulate the original functions into new versions F_i' such that the execution trace of all of them is some multiple of the execution trace of a shorter sequence. We term this shorted sequence Π, the fixed pattern of operations which is repeated to compose the larger functions. Clearly we might need to add some dummy instructions to the original functions as well as reordering their instructions so that the pattern is followed. To allow for instruction reordering, we extend our problem definition to include information about the data dependencies between instructions within each function. We represent these dependencies as directed graphs.

Definition 2. *Given a set F as in Definition 1, let P be the list of pairs*

$$P = (F_1, G_1), (F_2, G_2), ..., (F_N, G_N)$$

where $G_i = (V_i, E_i)$ is a directed graph in which V_i and E_i are the associated sets of nodes and edges, respectively. Let $|V_i| = |F_i|$ and, to each instruction $F_i[j]$, associate node $v_j \in V_i$. Let E_i contain an edge from node v_j to node v_k if and only if executing instruction $F_i[j]$ before instruction $F_i[k]$ disrupts the normal data flow inside the function. We say that instruction $F_i[j]$ depends on instruction $F_i[k]$.

In general terms, given a straight-line function F_i described using three-address operations from our instruction set L, the pair of function and graph (F_i,G_i) can be constructed as follows:

1. Add $|F_i|$ nodes to V_i so that each instruction in the function is represented by a node in the graph.
2. For every instruction $F_i[j]$ add an edge (v_j, v_k) to E_i if and only if $F_i[j]$ uses a result directly modified by some instruction $F_i[k]$. Note that we assume that symbols for intermediate results are not reused; that is the function is in single-static-assignment (SSA) form [19]. If reuse is permitted, additional edges must be inserted in the dependency graph to prevent overwriting intermediate results.
3. Calculate (V_i, E'_i), the transitive closure of the graph (V_i, E_i), and take $G_i = (V_i, E'_i)$.

We use the dependency graphs in Definition 2 to guarantee that the transformations we perform on the functions F_i are sound. That is, as long as we respect the dependencies, the program is functionally correct even though the instructions are reordered. Definition 3 captures this notion.

Definition 3. *A function F'_i is a valid transformation of a function F_i (written $F'_i \Lleftarrow F_i$) if given the dependency graph G_i, F'_i can be generated by modifying F_i as follows:*

1. *Reorder the instructions in F_i, respecting the dependency graph G_i i.e. if there is an edge $(v_j, v_k) \in E_i$ then instruction $F_i[j]$ must occur after instruction $F_i[k]$ in F'_i.*
2. *Insert a finite number of dummy instructions.*

The goal is to find Π and matching F'_i whose processing overhead compared to the original programs is minimised. Hence, our problem definition must also include the concept of computational cost. For the sake of generality, we assign to each basic instruction in set L an integer weight value that provides a relative measure of it's computational weight.

Definition 4. *Let $\omega : L \rightarrow \mathbb{N}$ be a weight function that, for each basic instruction $l \in L$, provides a relative value $\omega(l)$ for the computational load associated with instruction l.*

Given this cost function, we are now in a position to provide a formulation of the problem of building indistinguishable functions as an optimisation problem.

Definition 5. *Given a pair (P, ω) as in Definitions 1, 2 and 4, find a pattern Π and a list of functions $F' = F'_1, F'_2, ..., F'_N$ such that*

$$\begin{cases} \Pi = \Pi[1], \Pi[2], ..., \Pi[|\Pi|] & \Pi[k] \in L, 1 \leq k \leq |\Pi| \\ F'_i \Lleftarrow F_i & 1 \leq i \leq N \\ |F'_i| = 0 \pmod{|\Pi|} & 1 \leq i \leq N \\ F'_i[j] = \Pi[(j \bmod |\Pi|) + 1] & 1 \leq i \leq N, 1 \leq j \leq |F'_i| \end{cases}$$

and that

$$\sum_{i=1}^{N} \sum_{j=1}^{|F'_i|} \omega(F'_i[j])$$

is minimal.

To reiterate, from this definition we have that each function must composed of a number of instances of the pattern which constrains the type of each instruction. As a consequence, each instruction within each function matches the same instruction, modulo the pattern size, of every other function. Two functions are hence indistinguishable since one cannot identify their boundaries within a larger sequence of such patterns. In context, the only leaked information is potentially the Hamming weight and length of d: this is undesirable but unavoidable given the scope of our work.

Intuition on the hardness of satisfying these constraints comes from noticing similarities with well-known NP-complete optimisation problems such as the Minimum Bin Packing, Longest Common Subsequence and Nearest Codeword problems [9].

2.1 A Small Example

Recalling our definition of the elliptic curve $E(K)$ in Section 1.1, Algorithm 2 details two functions for affine point addition and doubling on such a curve. Denoting the addition and doubling as functions F_1 and F_2 respectively, we find $|F_1| = 10$ while $|F_2| = 13$. From these functions, we also find our instruction set is $L = \{x + y, x - y, x^2, x \times y, 1/x\}$ with all operations over the base field $K = \mathbb{F}_p$. Thus, we setup our costs as $\omega(x + y) = 1$, $\omega(x - y) = 1$, $\omega(x^2) = 10$, $\omega(x \times y) = 20$ and $\omega(1/x) = 100$.

Notice the role of dependencies in the functions: operation three in F_1 depends on operation two but not on operation one. In fact, we can relocate operation one after operation three to form a valid function F'_1 since it respects the data dependencies that exist.

The graphs in Figure 1 represent the direct dependencies between the instructions in the addition method (top) and the doubling method (bottom). Complete dependency graphs as specified in Definition 2 can be obtained by calculating the transitive closure over the graphs in Figure 1.

Algorithm 2 Methods for ECC affine point addition (left) and doubling (right).

```
Input: $P = (x_1, y_1), Q = (x_2, y_2)$
Output: $R = (x_3, y_3) = P + Q$
 1: $\lambda_1 \leftarrow y_2 - y_1$
 2: $\lambda_2 \leftarrow x_2 - x_1$
 3: $\lambda_3 \leftarrow \lambda_2^{-1}$
 4: $\lambda_4 \leftarrow \lambda_1 \cdot \lambda_3$
 5: $\lambda_5 \leftarrow \lambda_4^2$
 6: $\lambda_6 \leftarrow \lambda_5 - x_1$
 7: $x_3 \leftarrow \lambda_6 - x_2$
 8: $\lambda_7 \leftarrow x_1 - x_3$
 9: $\lambda_8 \leftarrow \lambda_4 \cdot \lambda_7$
10: $y_3 \leftarrow \lambda_8 - y_1$
```

```
Input: $P = (x_1, y_1)$
Output: $R = (x_3, y_3) = 2 \cdot P$
 1: $\lambda_1 \leftarrow x_1^2$
 2: $\lambda_2 \leftarrow \lambda_1 + \lambda_1$
 3: $\lambda_3 \leftarrow \lambda_2 + \lambda_1$
 4: $\lambda_4 \leftarrow \lambda_3 + A$
 5: $\lambda_5 \leftarrow y_1 + y_1$
 6: $\lambda_6 \leftarrow \lambda_5^{-1}$
 7: $\lambda_7 \leftarrow \lambda_4 \cdot \lambda_6$
 8: $\lambda_8 \leftarrow \lambda_7^2$
 9: $\lambda_9 \leftarrow x_1 + x_1$
10: $x_3 \leftarrow \lambda_8 - \lambda_9$
11: $\lambda_{10} \leftarrow x_1 - x_3$
12: $\lambda_{11} \leftarrow \lambda_{10} \cdot \lambda_7$
13: $y_3 \leftarrow \lambda_{11} - y_1$
```

Algorithm 3 shows a solution for this instance of the optimisation problem. The cost of the solution is 12 since we add an extra square and two extra additions both denoted by the use of λ_d as their arguments. It is easy to see that it is actually an absolute minimal value. To clarify the criteria specified in Definition 5, let us see how they apply to this case.

The pattern Π is given by the operation sequence of the doubling method, and we have $|\Pi| = 13$. To ensure both $|F'_1| = 0 \pmod{|\Pi|}$ and $|F'_2| = 0 \pmod{|\Pi|}$ we need to add three dummy instructions to F_1. The solution presents no mismatches between the instruction sequences of either function and the pattern Π, so the restriction $F'_i[j] = \Pi[(j \bmod |\Pi|) + 1]$ holds for all valid i and j values. Finally, it is easy to see that both F'_i are valid transformations of the original F_i. Instruction reordering occurs only once in F'_1 (instructions 3 and 5), and these are independent in F_1. F'_2 is identical to F_2.

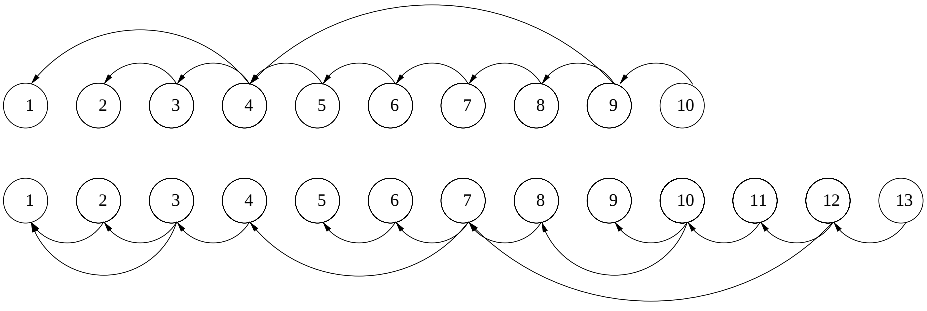

Fig. 1. Dependency graphs for the methods in Algorithm 2.

Algorithm 3 Indistinguishable versions of the methods in Algorithm 2.

```
Input: $P = (x_1, y_1), Q = (x_2, y_2)$
Output: $R = (x_3, y_3) = P + Q$
$\lambda_d \leftarrow \lambda_d^2$
$\lambda_d \leftarrow \lambda_d + \lambda_d$
$\lambda_2 \leftarrow x_2 - x_1$
$\lambda_d \leftarrow \lambda_d + \lambda_d$
$\lambda_1 \leftarrow y_2 - y_1$
$\lambda_3 \leftarrow \lambda_2^{-1}$
$\lambda_4 \leftarrow \lambda_1 \cdot \lambda_3$
$\lambda_5 \leftarrow \lambda_4^2$
$\lambda_6 \leftarrow \lambda_5 - x_1$
$x_3 \leftarrow \lambda_6 - x_2$
$\lambda_7 \leftarrow x_1 - x_3$
$\lambda_8 \leftarrow \lambda_4 \cdot \lambda_7$
$y_3 \leftarrow \lambda_8 - y_1$
```

```
Input: $P = (x_1, y_1)$
Output: $R = (x_3, y_3) = 2 \cdot P$
$\lambda_1 \leftarrow x_1^2$
$\lambda_2 \leftarrow \lambda_1 + \lambda_1$
$\lambda_3 \leftarrow \lambda_2 + \lambda_1$
$\lambda_4 \leftarrow \lambda_3 + A$
$\lambda_5 \leftarrow y_1 + y_1$
$\lambda_6 \leftarrow \lambda_5^{-1}$
$\lambda_7 \leftarrow \lambda_4 \cdot \lambda_6$
$\lambda_8 \leftarrow \lambda_7^2$
$\lambda_9 \leftarrow x_1 + x_1$
$x_3 \leftarrow \lambda_8 - \lambda_9$
$\lambda_{10} \leftarrow x_1 - x_3$
$\lambda_{11} \leftarrow \lambda_{10} \cdot \lambda_7$
$y_3 \leftarrow \lambda_{11} - y_1$
```

3 An Optimisation Algorithm

Our approach to solving the problem as described above is detailed in Algorithm 4. The algorithm represents an adaptation of Threshold Accepting [10], a generic optimisation algorithm. Threshold Accepting is a close relative of simulated annealing, where candidate solutions are deterministically accepted or rejected according to a predefined threshold schedule: a proposed solution is accepted as long as it does not increase the current cost by more than the current threshold value. Note that we are not aiming to find the optimal solution, but to find a good enough approximation of it that can be used in practical applications.

Algorithm 4 makes S attempts to find an optimal pattern size, which is selected randomly in each s-iteration (line 2). In each of these attempts, the original functions are taken as the starting solution, with the minor change that they are padded with dummy instructions, so as to make their size multiple of the pattern size (lines 3 to 6).

The inner loop (lines 10 to 18), which runs T times, uses a set of randomised heuristics to obtain a neighbour solution. This solution is accepted if it does not represent a relative cost increase greater than the current threshold value. The threshold varies with t, starting at a larger value for low values of t and gradually decreasing. The number of iterations S and T must be adjusted according to the size of the problem.

The quality of a solution is measured using a cost function that operates as follows:

- The complete set of instructions in a solution $\mathbf{x}$ is seen as a matrix with $|\Pi|$ columns and $(\sum_{i=1}^{N} |F'_i|)/|\Pi|$ rows (see Figure 2), in which each function occupies $|F'_i|/|\Pi|$ consecutive rows.

Algorithm 4 An optimisation algorithm for indistinguishable operations.

```
Input: $(\Pi, \omega)$
Output: $(\Pi, F')$, a quasi-optimal solution to the problem in Definition 5
 1: for $s = 1$ to $S$ do
 2:   $|\Pi| \leftarrow$ random pattern size
 3:   $\mathbf{x} \leftarrow \{F_i, 1 \leq i \leq N\}$
 4:   for all $F'_i \in \mathbf{x}$ do
 5:     Append $|\Pi| - (|F'_i| \pmod{|\Pi|})$ dummy instructions to $F'_i$
 6:   end for
 7:   $cost \leftarrow cost(\mathbf{x})$
 8:   $result \leftarrow \mathbf{x}$
 9:   $best \leftarrow cost$
10:   for $t = 1$ to $T$ do
11:     $\mathbf{x}' \leftarrow neighbour(\mathbf{x})$
12:     $thresh \leftarrow threshold(t, T)$
13:     $cost' \leftarrow cost(\mathbf{x}')$
14:     if $(cost' - cost)/cost - 1 < thresh$ then
15:       $\mathbf{x} \leftarrow \mathbf{x}'$
16:       $cost \leftarrow cost'$
17:     end if
18:   end for
19:   if $cost < best$ then
20:     $result \leftarrow \mathbf{x}$
21:     $best \leftarrow cost$
22:   end if
23: end for
24: return $result$
```

- Throughout the algorithm, the pattern Π is adjusted to each solution by taking $\Pi[k]$ as the most common basic instruction in column k of the matrix.
- A dummy instruction is always taken to be of the same type as the pattern instruction for its particular column, so dummy instructions are ignored when adjusting Π to a particular solution.
- The overall cost of a solution has two components: c and d. The former is the cost associated with deviations from the pattern and it is evaluated as the sum, taken over all non-matching instructions in the matrix, of the weight difference relative to the corresponding pattern instruction. The latter is the cost associated with dummy operations and it is evaluated as the accumulated weight of all the dummy instructions in the matrix.
- The relative importance of these components can be tuned to put more or less emphasis on indistinguishability. This affects the trade-off between indistinguishability and processing overhead.

Throughout its execution, the algorithm keeps track of the best solution it has been able to reach (lines 19 to 22); this is returned when the algorithm terminates (line 24).

$$\begin{array}{ccccc}
\mathit{\Pi}[1] & \mathit{\Pi}[2] & ... & \mathit{\Pi}[|\mathit{\Pi}|-1] & \mathit{\Pi}[|\mathit{\Pi}|] \\
\left[\begin{array}{c} F_1'[1] \\ F_1'[|\mathit{\Pi}|+1] \\ \vdots \\ F_1'[|F_1'|-|\mathit{\Pi}|+1] \\ F_2'[1] \\ \vdots \\ F_N'[|F_N'|-|\mathit{\Pi}|+1] \end{array}\right.
& \begin{array}{c} F_1'[2] \\ F_1'[2] \\ \vdots \\ F_1'[|F_1'|-|\mathit{\Pi}|+2] \\ F_2'[2] \\ \vdots \\ F_N'[|F_N'|-|\mathit{\Pi}|+2] \end{array}
& \begin{array}{c} ... \\ ... \\ \vdots \\ ... \\ ... \\ \vdots \\ ... \end{array}
& \begin{array}{c} F_1'[|\mathit{\Pi}|-1] \\ F_1'[2|\mathit{\Pi}|-1] \\ \vdots \\ F_1'[|F_1'|-1] \\ F_2'[|\mathit{\Pi}|-1] \\ \vdots \\ F_N'[|F_N'|-1] \end{array}
& \left.\begin{array}{c} F_1'[|\mathit{\Pi}|] \\ F_1'[2|\mathit{\Pi}|] \\ \vdots \\ F_1'[|F_1'|] \\ F_2'[|\mathit{\Pi}|] \\ \vdots \\ F_N'[|F_N'|] \end{array}\right]
\end{array}$$

Fig. 2. A solution as a matrix of instructions.

A neighbour solution is derived from the current solution by randomly selecting one of the following heuristics:

Tilt function left A function F_i' is selected randomly, and its instructions are all moved as far to the left as possible, filling spaces previously occupied by dummy instructions or just freed by moving other instructions. The order of the instructions is preserved, and an instruction is only moved if it matches the pattern instruction for the target column.

Tilt function right Same as previous, only that instructions are shifted to the right.

Move instruction left A function F_i' is selected randomly, and an instruction $F_i'[j]$ is randomly picked within it. This instruction is then moved as far to the left as possible. An instruction is only moved if this does not violate inter-instruction dependencies, and it matches the pattern instruction for the target column.

Move instruction right Same as previous, only that the instruction is shifted to the right.

After application of the selection heuristic, the neighbour solution is optimised by removing rows or columns containing only dummy instructions. If the final solution produced by Algorithm 4 includes deviations from the chosen pattern, these can be eliminated in an optional post-processing phase. In this phase we increase the pattern size to cover the mismatches and introduce extra dummy operations to retain indestinguishability. If the number of mismatches is large, this produces a degenerate solution which is discarded due to the high related cost.

Our experimental results indicate the following rules of thumb that should be considered when parameterising Algorithm 4:

- S should be at least half of the length of the longest function.
- T should be a small multiple of the total number of instructions.
- An overall cost function calculated as $c^2 + d$ leads to a good compromise between indistinguishability and processing overhead.

- The threshold should decrease quadratically, starting at 70% for $t = 0$ and reaching 10% when $t = T - 1$.
- The neighbour generation heuristics should be selected with equal probability, or with a small bias favouring moving over tilting mutations.

4 Results

Using Algorithm 4 we have been able to produce results equivalent to various hand-made solutions published in the literature for small sized problems, and to construct indistinguishable versions of the much larger functions for point addition and doubling in genus 2 hyper-elliptic curves over finite fields. To save space, we refer to Appendices within the full version of this paper for a complete set of results [3].

Appendix 1 contains the results produced by Algorithm 4 when fed with the instruction sequences for vanilla EC affine point addition over $\mathbb{F}_p$ using projective coordinates, as presented by Gebotys and Gebotys in (Figure 1,[12]). This result has exactly the same overhead as the version presented in the same reference. Appendix 2 contains the instruction sequences corresponding to formulae for finite field arithmetic in a specific degree six extension as used in a number of torus based constructions [13, 11], together with the results obtained using Algorithm 4 for this problem instance.

Table 1 shows the number of dummy field operations added to each of the functions in Appendix 2. Also shown in Table 1 is the estimated overhead for an exponentiation. We assume the average case in which the number of squarings is twice the number of multiplications. This is roughly equivalent to the best hand-made solution that we were able to produce in reasonable time, even if the number of dummy multiplications is slightly larger in the automated version.

Table 1. Overheads for the indistinguishable functions in Appendix 2.

Operations	*Square*	*Multiply*	Overhead
Add	9	11	24%
Multiply	4	2	19%
Shift	0	12	400%

Appendix 3 includes indistinguishable versions of hyper-elliptic curve point adding and doubling functions. These implementations correspond to the general case of the explicit formulae for genus 2 hyperelliptic curves over finite fields using affine coordinates. provided by Lange in [17]. A pseudo-code implementation of these formulae is also included in Appendix 3. In our analysis, we made no assumptions as to the values of curve parameters because our objective was to work over a relatively large problem instance. Operations involving curve parameters were treated as generic operations.

In this example, the group operations themselves contain branching instructions. To accommodate this, we had to first create indistinguishable versions of the smaller branches inside the addition and doubling functions, separately. The process was then applied globally to the two main branches of the addition function and to the doubling function as a whole, which meant processing three functions of considerable size, simultaneously.

Table 2 (left) shows the number of dummy field operations added to each of the functions in Appendix 3. Note that functions $Add2'$ and $Double'$ correspond to cases that are overwhelmingly more likely to occur. The overhead, in these cases, is within reasonable limits. Also shown in Table 2 (right) is the estimated overhead for a point multiplication. We assume the average case in which the number of doublings is twice the number of additions, and consider only the most likely execution sequences for these operations ($Add2'$ and $Double'$).

Table 2. Overheads for the indistinguishable functions in Appendix 3.

Operations	$Add1$	$Add2'$	$Add2''$	$Double'$	$Double''$	Overhead
Add	35	20	37	4	16	19%
Multiply	28	14	27	5	16	25%
Square	7	5	6	2	4	60%
Invert	1	1	1	1	1	100%

5 Conclusion

Defence against side-channel attacks is a vital part of engineering software that is to be executed on constrained devices. Since such devices are used within an adversarial environment, sound and efficient techniques are of value even if they are hard to implement. To this end we have investigated an automated approach to constructing indistinguishable functions, a general method of defence against certain classes of side-channel attack which are notoriously difficult to implement as the functions grow more complex. Our results show that efficient versions of such functions, which are competitive with hand-constructed versions, can be generated with only minor computational effort.

This work is pitched in the context of cryptography-aware compilation: the idea that programmers should be assisted in describing secure software just like they are offered support to optimise software. We have embedded our algorithm in such a compiler which can now automatically manipulate a source program so the result is more secure. As such, interesting further work includes addressing the relationship between register allocation and construction of indistinguishable functions. Ideally, the number of registers used is minimised using, for example, a graph colouring allocator. Manipulation of functions can alter the effectiveness of this process, a fact that requires some further investigation. Equally, the relationship between the presented work and side-channel atomicity might provide an avenue for further work. One would expect a similar method to the one presented here to be suitable for automatic construction of side-channel atomic patterns, and that aggressive inlining within our compiler could present the opportunity to deploy such a method.

Acknowledgements

The authors would like to thank Nigel Smart and Richard Noad for their input during this work, and the anonymous reviewers whose comments helped to improve it.

References

1. D. Agrawal, B. Archambeault, J.R. Rao and P. Rohatgi. The EM Side-Channel(s). In *Cryptographic Hardware and Embedded Systems (CHES)*, Springer-Verlag LNCS 2523, 29–45, 2002.
2. D. Agrawal, J.R. Rao and P. Rohatgi. Multi-channel Attacks. In *Cryptographic Hardware and Embedded Systems (CHES)*, Springer-Verlag LNCS 2779, 2–16, 2003.
3. M. Barbosa and D. Page. On the Automatic Construction of Indistinguishable Operations. In *Cryptology ePrint Archive*, Report 2005/174, 2005.
4. I.F. Blake, G. Seroussi and N.P. Smart. *Elliptic Curves in Cryptography.* Cambridge University Press, 1999.
5. I.F. Blake, G. Seroussi and N.P. Smart. *Advances in Elliptic Curve Cryptography.* Cambridge University Press, 2004.
6. É. Brier and M. Joye. Weierstraß Elliptic Curves and Side-channel Attacks. In *Public Key Cryptography (PKC)*, Springer-Verlag LNCS 2274, 335–345, 2002.
7. B. Chevallier-Mames, M. Ciet and M. Joye. Low-Cost Solutions for Preventing Simple Side-Channel Analysis: Side-Channel Atomicity. In *IEEE Transactions on Computers*, 53(6), 760–768, 2004.
8. J-S. Coron. Resistance against Differential Power Analysis for Elliptic Curve Cryptosystems. In *Cryptographic Hardware and Embedded Systems (CHES)*, Springer-Verlag LNCS 1717, 292–302, 1999.
9. P. Crescenzi and V. Kann. A Compendium of NP Optimization Problems. Available at: http://www.nada.kth.se/∼viggo/problemlist/.
10. G. Dueck and T. Scheuer. Threshold Accepting: A General Purpose Optimization Algorithm Appearing Superior to Simulated Annealing. In *Journal of Computational Physics*, 90(1), 161–175, 1990.
11. M. van Dijk, R. Granger, D. Page, K. Rubin, A. Silverberg, M. Stam and D. Woodruff. Practical Cryptography in High Dimensional Tori. *Advances in Cryptology (EUROCRYPT)*, Springer-Verlag LNCS 3494, 234–250, 2005.
12. C.H. Gebotys and R.J. Gebotys. Secure Elliptic Curve Implementations: An Analysis of Resistance to Power-Attacks in a DSP Processor. In *Cryptographic Hardware and Embedded Systems (CHES)*, Springer-Verlag LNCS 2523, 114–128, 2002.
13. R. Granger, D. Page and M. Stam. A Comparison of CEILIDH and XTR. In *Algorithmic Number Theory Symposium (ANTS)*, Springer-Verlag LNCS 3076, 235–249, 2004.
14. M. Joye and J-J. Quisquater. Hessian Elliptic Curves and Side-Channel Attacks. In *Cryptographic Hardware and Embedded Systems (CHES)*, Springer-Verlag LNCS 2162, 402–410, 2001.

15. P.C. Kocher. Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems. In *Advances in Cryptology (CRYPTO)*, Springer-Verlag LNCS 1109, 104–113, 1996.
16. P.C. Kocher, J. Jaffe and B. Jun. Differential Power Analysis. In *Advances in Cryptology (CRYPTO)*, Springer-Verlag LNCS 1666, 388–397, 1999.
17. T. Lange. Efficient Arithmetic on Genus 2 Hyperelliptic Curves over Finite Fields via Explicit Formulae. In *Cryptology ePrint Archive*, Report 2002/121, 2002.
18. P-Y. Liardet and N.P. Smart. Preventing SPA/DPA in ECC Systems Using the Jacobi Form. In *Cryptographic Hardware and Embedded Systems (CHES)*, Springer-Verlag LNCS 2162, 391–401, 2001.
19. S.S. Muchnick. *Advanced Compiler Design and Implementation*, Morgan Kaufmann, 1997.
20. D. Page and M. Stam. On XTR and Side-Channel Analysis. In *Selected Areas in Cryptography (SAC)*, Springer-Verlag LNCS 3357, 54–68, 2004.
21. E. Trichina and A. Bellezza. Implementation of Elliptic Curve Cryptography with Built-In Counter Measures against Side Channel Attacks. In *Cryptographic Hardware and Embedded Systems (CHES)*, Springer-Verlag LNCS 2523, 98–113, 2002.