Modelling the Impact of Code Obfuscation on Energy Usage

Athul Raj, Jithish J and Sriram Sankaran
Amrita Center for Cybersecurity Systems and Networks
Amrita School of Engineering, Amritapuri
Amrita Vishwa Vidyapeetham
Amrita University
India
srirams@am.amrita.edu

ABSTRACT
Advancements in computing and communication technologies have given rise to low-cost embedded devices with applications in diverse domains such as Smarthome, industrial automation, healthcare, transportation etc. These devices are power-constrained which emphasizes the need for lightweight security solutions. Code obfuscation has been demonstrated to provide time-limited protection of source code from inference or tampering attacks. However, size of the obfuscated code increases with increase in code size which can have a negative impact on energy consumption. In particular, different transformations of the source code result in varying amounts of energy consumption for embedded devices. In this work, we model the impact of algorithms for code obfuscation on energy usage for embedded devices and analyze the energy-security-performance trade-offs. The insights from our analysis can be used to develop techniques depending on the needs of the applications thus facilitating efficient energy usage.

1. INTRODUCTION
The proliferation of low-cost embedded devices with advancements in computing and communication capabilities facilitates device-to-device communication with varying capabilities. Energy Management has become one of the foremost concerns in mobile devices. This is primarily due to the increasing functionality in applications which rapidly drains battery power in these devices. In addition, security contributes towards energy drain due to the significant performance overhead of mechanisms used to secure embedded devices. These emphasize the need for developing energy aware security mechanisms for embedded devices. Further these mechanisms require measurement tools for estimating the energy consumed due to security for embedded devices.

Due to the increasing sensitivity of software applications stored on embedded devices, mechanisms are necessary for protecting their source code which runs on these devices. Code obfuscation has been shown to provide time-limited protection of source codes thus preventing intruders from tampering or inference attacks. The primary role of a code obfuscator is to apply a transformation to the original source code thus resulting in obfuscated source code which in turn is stored in the embedded devices. This results in obstructing the regular control flow and performing software manipulations thus making it impossible to reverse engineer the source code. In addition, code obfuscation has a considerable impact on performance and energy consumption which varies for different kinds of applications.

While code obfuscation has been demonstrated to provide time-limited protection for embedded devices, its impact on energy consumption needs to be analyzed. In particular, increase in size of the source code results in corresponding increase in size of the obfuscated code thus causing a negative impact on power consumption. Further different transformations of the source code result in varying amounts of energy consumption for embedded devices. Thus a comprehensive energy analysis of algorithms for code obfuscation is necessary to analyze the energy-performance trade-offs. In addition, estimating security is necessary to analyze the effectiveness of the obfuscation techniques. However, estimating security of obfuscated code is challenging due to the varying number of transformations applied to source codes with increasing functionality.

In this work, we model the impact of algorithms for code obfuscation on energy usage in embedded devices. In particular, we study different techniques for obfuscation such as Lexical, Data and Control obfuscation and analyze their impact on energy usage. We describe the experiment set-up for measuring energy and performance and present our analysis of energy-performance trade-offs. Our analysis conducted using Mibench benchmarks reveals that data and control flow obfuscation incur significant energy consumption compared to lexical obfuscation. The results obtained from this study can be used to tailor obfuscation to suit the needs of resource-constrained devices and further analyze energy-security-performance trade-offs.

2. RELATED WORK
Numerous approaches have studied the problem of code obfuscation and analyzed their applicability in embedded devices. Collberg et. al [9] presented a taxonomy of different kinds of obfuscation which describes transformations for
securing diverse source codes. Further, there exists tools for code obfuscation at both the hardware and software levels. While some of the tools such as Tigress [8] and ProGuard [18] are open-source, there exists numerous commercial tools such as Allatori [1], Dasho [2], Zelix [5].

There exists approaches for measuring the impact of obfuscated code on energy consumption. In particular Sahin et al. [21] profiled the energy consumption of different transformations on the Android smartphone and statistically analyzed the significance between normal and obfuscated code. However, they claimed that energy consumption between different methods of a particular kind of obfuscation is insignificant. In addition, Dukovic et al. [11] profiled the energy consumption of normal and obfuscated code at an instruction-level and analyzed the impact of different transformations.

In addition to profiling energy consumption, certain approaches have analyzed the security of obfuscated code for diverse kinds of systems. Banescu et al. [6] measured the resilience of obfuscated code against reverse engineering using tools such as tiger which contains numerous transformations. Wu et al. [26] estimated the security of different obfuscation based techniques using approximation. In particular, their approach involves modelling the relationship between obfuscation parameters and the corresponding impact on security using regression-based techniques.

While energy and security are critical, performance is equally necessary as to meet real-time constraints for embedded applications. This can be achieved using hardware obfuscation which is necessary to protect secret information in applications. This can be achieved using hardware obfuscation based techniques using approximation. In particular, their approach involves modelling the relationship between obfuscation parameters and the corresponding impact on security using regression-based techniques.

While energy consumption of computing systems in general and embedded devices. Economou et al. [12] and Isci et al. [15] modeled the energy consumption of embedded devices using performance counters. Khan et al. [17] and Sankaran et al. [23] modeled the energy consumption of multi-core systems using a statistical learning approach. Wang et al. [25] developed SPAN, a software power monitoring tool which correlates program segments with power consumption. Fan et al. [13] analyzed the power consumption characteristics of data centers and studied the optimal provisioning of resources. Sangaiah et al. [22] developed regression models to predict the performance of Chip Multiprocessors.

In contrast to the existing approaches, we profile the energy consumption and performance of algorithms for code obfuscation in embedded devices. Although our work is closely related to [11] and [21], we manually obfuscate source code of benchmark applications which enables us to explore numerous transformations in contrast to the obfuscation tools used by [11] and [21]. While tools for obfuscation are available, they are often commerical and that we are limited by the transformations available in those tools. In addition, our experiments are conducted using SourceMeter, a power measurement tool which provides accurate power estimates.

3. BACKGROUND

3.1 Lexical Obfuscation

Lexical obfuscation is one of the basic and simplest form of obfuscation used in software programs. It includes a wide array of operations such as comment removal, identifier renaming, structured construction removal, debugging info removal etc. Typically lexical obfuscation is used for identifier renaming which may not have any impact on the security of the source code. Thus it becomes relatively easy for the attacker to understand and reverse-engineer the source code. In addition, other techniques such as structured construction removal, debugging info removal and comment removal may reduce the size of the source code. While lexical obfuscation lacks security, it has been either replaced or used in conjunction with other techniques such as data and control obfuscation.

3.2 Data Obfuscation

Data obfuscation provides stronger security than lexical obfuscation by protecting data in the source code. In particular, data is protected in such a way such that it becomes hard to infer the functionality through code analysis. Data obfuscation includes a wide array of operations such as string scrambling, array restructuring and merging, data encoding and variable reordering. For instance, value of a variable can be changed to include numerous different variables. Thus the value of the variable can be determined by fusing the contents of the created variables. Similarly, arrays can be restructured by creating arbitrary number of new arrays and merging them with existing ones. Techniques for data obfuscation have been described in [10] and [20]. Thus by combining different kinds of techniques for data obfuscation, overall security can be enhanced. However, size of the obfuscated code may increase which causes a negative impact on power and execution time.

3.3 Control Obfuscation

Control obfuscation obfuscates the control flow of the program thus providing stronger security than lexical and data obfuscation. This includes manipulating the flow of execution with irrelevant conditional statements which in turn results in restructuring of methods, loops and statements. Typically, restructuring consists in inlining, outlining, inter-leaving and cloning of functions and elimination of library calls [24] [7]. In addition, a single function is transformed through opaque predicates, insertion of dead code etc. However, as the control flow of the program is obfuscated, it becomes harder for the attacker to interrelate the various sections of the program. As a result, there is an increase in the size of the obfuscated code which negatively impacts power consumption and execution time.

4. EXPERIMENT SET-UP

In this section, we discuss the hardware and software used for our experiments along with the description of our power measurement set-up for energy analysis. We use the STM32-F0DISCOVERY [19] development board from STMMicroelectronics as the embedded platform for measuring the energy impact of code obfuscation. The board operates at a DC power supply of 3V. The board consists of an ARM Cortex-M0 microcontroller with 64KB flash and 8KB RAM. In addition, Keil MDK-ARM embedded software development envi-
For evaluating different obfuscation techniques.

**Power Measurement:**
Keithley’s Series 2400 Source Measure Unit (SMU) [4] is used for measuring the power consumption of the board. It integrates both source and measurement circuitry in a single unit thus facilitating a fast and accurate measurement of power consumption. Figure 1 contains the pictorial representation of the set-up used for our experiments.

The SMU is used as the DC power source for the set-up. It provides the required stable precision DC power supply of 3V for the embedded board. The obfuscated programs are developed on the Keil MDK-ARM software suite running on Windows 10 platform and flashed to the board via USB interface. The SMU logs the power consumed by the embedded board at distinct time intervals. Further, we perform comparative analysis of power traces using MATLAB.

**Benchmark Applications:**
We consider the following applications from the MiBench benchmark suite [14] for comparatively analyzing the impact of code obfuscation on energy usage. Similarly other applications can be profiled and its energy impact on code obfuscation analyzed.

- **Basicmath:** Mathematical calculations involving cubic function solving, square root evaluation, degrees to radian conversion etc. on a fixed set of constants.
- **Bitcount:** Evaluating bit manipulation capabilities by computing the total number of bits in a fixed input array of integers.
- **Matrix Multiplication:** Computing the matrix multiplication which is of complexity \(O(n^3)\).
- **Vernam Cipher:** Symmetric stream cipher which utilizes Boolean XOR operation to generate the ciphertext.

## 5. ENERGY ANALYSIS

In this section, we analyze the energy consumption of different code obfuscation techniques using the embedded benchmarks. In particular, we consider obfuscation techniques such as lexical, data and control flow transformations. Within the context of data obfuscation, we transform array data [20] for code obfuscation. In control flow obfuscation, we perform control flow flattening, dead code insertion, extended loop condition and loop transformation. Further, we analyze a transformation which involves a combination of above obfuscation techniques termed as combined obfuscation.

In our experiments, we manually obfuscated the source code of the applications using the above transformations. While there exists open-source and commercial tools for code obfuscation, they are limited in terms of the number of available transformations. Thus, our goal is to explore numerous transformations and their possible combinations and not be restricted by the tools.

Our analysis reveals that different operations performed by the Cortex M0 microcontroller have corresponding power consumption profiles. Thus, our goal is to analyze the impact of different obfuscation techniques on the energy and performance of embedded devices. Towards this goal, we gather the power traces from the SMU for normal and obfuscated codes pertaining to different transformations. Further, execution time and energy consumption were estimated for each of the applications. In particular, we compute the percentage difference of parameters such as average power, execution time, energy consumption and storage between normal and obfuscated code.

Table 1 contains the power, execution time, energy and storage results for lexical, data and combined obfuscation while table 2 displays those for each of the transformations in control flow obfuscation. From the table, \(\%\Delta P, \%\Delta T, \%\Delta E\) and \(\%\Delta S\) denote the percentage difference of parameters such as power consumption, execution time, energy and storage respectively between normal and obfuscated code.

From the tables, it is evident that the change in energy consumption ranges from 0% to 11.32% for different obfuscation techniques and that the maximum change is observed in basicmath when data obfuscation is applied. In contrast, minimum change is observed for matrix multiplication and Vernam cipher when data and lexical obfuscation are applied respectively. In the following, we analyze the energy
Table 2: Control Flow Obfuscation

<table>
<thead>
<tr>
<th>App</th>
<th>%ΔP</th>
<th>%ΔT</th>
<th>%ΔE</th>
<th>%ΔS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matrix</td>
<td>CF:1.27</td>
<td>CF:0</td>
<td>CF:1.27</td>
<td>CF:3.76</td>
</tr>
<tr>
<td></td>
<td>DI:2.17</td>
<td>DI:4</td>
<td>DI:6.26</td>
<td>DI:8.85</td>
</tr>
<tr>
<td></td>
<td>EL:2.35</td>
<td>EL:4</td>
<td>EL:6.45</td>
<td>EL:4.5</td>
</tr>
<tr>
<td></td>
<td>LT:1.87</td>
<td>LT:4</td>
<td>LT:5.95</td>
<td>LT:9.01</td>
</tr>
<tr>
<td>Basicmath</td>
<td>CF:3.62</td>
<td>CF:5.88</td>
<td>CF:9.71</td>
<td>CF:5.57</td>
</tr>
<tr>
<td></td>
<td>DI:0.95</td>
<td>DI:5.88</td>
<td>DI:6.89</td>
<td>DI:1.18</td>
</tr>
<tr>
<td></td>
<td>EL:1.21</td>
<td>EL:5.88</td>
<td>EL:7.16</td>
<td>EL:1.58</td>
</tr>
<tr>
<td></td>
<td>LT:0.38</td>
<td>LT:2.94</td>
<td>LT:3.33</td>
<td>LT:0.57</td>
</tr>
<tr>
<td>Bitcount</td>
<td>CF:1.18</td>
<td>CF:1.4</td>
<td>CF:2.6</td>
<td>CF:3.59</td>
</tr>
<tr>
<td></td>
<td>DI:1.8</td>
<td>DI:1.41</td>
<td>DI:3.24</td>
<td>DI:3.83</td>
</tr>
<tr>
<td></td>
<td>EL:1.37</td>
<td>EL:1.41</td>
<td>EL:1.37</td>
<td>EL:4.43</td>
</tr>
<tr>
<td></td>
<td>LT:0.12</td>
<td>LT:1.41</td>
<td>LT:0.12</td>
<td>LT:1.02</td>
</tr>
<tr>
<td>Vernam</td>
<td>CF:0.5</td>
<td>CF:8.33</td>
<td>CF:8.88</td>
<td>CF:3.13</td>
</tr>
<tr>
<td></td>
<td>DI:0.82</td>
<td>DI:8.33</td>
<td>DI:9.22</td>
<td>DI:2.99</td>
</tr>
<tr>
<td></td>
<td>EL:1.01</td>
<td>EL:8.33</td>
<td>EL:9.43</td>
<td>EL:3.74</td>
</tr>
<tr>
<td></td>
<td>LT:0.25</td>
<td>LT:8.33</td>
<td>LT:8.61</td>
<td>LT:2.51</td>
</tr>
</tbody>
</table>

CF: Control Flow Flattening  DI: Deadcode Insertion  EL: Extended Loop Condition  LT: Loop Transformation

and performance impact pertaining to each of the obfuscation techniques.

**Lexical Obfuscation:**

The least values for change in energy consumption across all applications indicate that lexical obfuscation minimally impacts power consumption. This is due to the negligible overhead involved in carrying out the operations such as changing variables, removing comments etc. Thus overall structure of the actual code is maintained with minimal changes.

Similarly, we observe that execution time between normal and obfuscated code remains similar for the four applications. Thus lexical obfuscation incurs minimal impact on performance.

**Data Obfuscation:**

In the case of data obfuscation, we observe a noticeable increase in energy consumption. In particular, data obfuscation of basicmath consumes higher energy compared to other transformations. This is due to the data manipulation operations such as cubic function solving, square root evaluation, degree to radian conversion etc that are contained in data obfuscation which incurs significant amount of computation resulting in high energy.

Similarly, we observe a maximum increase in execution time of 8.82 % for basicmath program followed by Vernam cipher of 4.16% and close to negligible for matrix multiplication and bitcount benchmarks. Thus data obfuscation negatively impacts performance of basicmath due to the compute intensive operations.

**Control Flow Obfuscation:**

Among the four transformations for control flow obfuscation, we observe maximum energy consumption for extended loop condition and loop transformations. This is primarily due to the increase in number of loop operations which may be computationally intensive and that they consume more energy compared to other transformations.

Similarly, we observe a noticeable increase in execution time for the basicmath and Vernam cipher benchmarks due to control obfuscation. This can be attributed to the increase in the number of data manipulation operations compared to other benchmark applications.

**Combined Obfuscation**

From the tables, it is evident that combined obfuscation incurs a maximum impact on energy consumption. In particular we observe a maximum increase in energy consumption of 13.8% for Vernam cipher followed by basicmath and bitcount benchmarks that are at 13.32% and 13.8% respectively. This is primarily due to the greater code size after combining the techniques of obfuscation which resulted in high energy consumption.

Similarly, combined obfuscation negatively impacts performance by increasing the execution time of obfuscated programs due to combined obfuscation.

### 5.1 Trace Analysis

To analyze the long-term impact of code obfuscation, we gather power traces using SMU for different obfuscation techniques such as lexical, data and control flow transformations. Figures 2, 3 and 4 pictorially represent the results for lexical, data and control flow transformations pertaining to each of the applications. From the figure, it is evident that for certain transformations such as lexical obfuscation, power traces are similar and variations are minimal. To examine the degree of correlation between traces, we utilize a statistical measure called Pearson correlation coefficient.

Pearson correlation coefficient is a statistical measure used to estimate the degree of similarity between different kinds of data. A value of 1.00 implies perfect correlation, while 0 and -1 indicate no-correlation and negative correlation respectively. We estimate the correlation coefficient between obfuscated code and non-obfuscated code for the transformations pertaining to each of the applications. Table 3 contains the results for correlation.

From the table, we make the following inferences. Lexical obfuscation maintains the structure information of the actual code without changing the original assembly code. This explains the similarity in cross correlation values for normal and lexically obfuscated code. In the case of data obfuscation, it is evident from figure 3 that the power traces are not similar. This behavior is validated by the cross correlation values computed for individual power traces. Similarly, transformations that involve loop operations such as extended loop and loop transformations exhibit wide variations in control flow obfuscation. However, we observed a
maximum variation in traces in the case of combined obfuscation. This is further validated by the relatively least correlation coefficient for these transformations.

Figure 3: Data Obfuscation

Figure 4: Combined Obfuscation

6. ENERGY OPTIMIZATION

Our study shows that obfuscation contributes significantly towards the energy consumed by embedded devices and that the rate of energy consumed depends on the kind of obfuscation. Thus, optimizing the energy consumed due to obfuscation is necessary so that systems meet a given power budget. As mobile applications are increasingly becoming critical and that the power consumed due to securing them may proportionately increase, minimizing energy use is critical considering the long-term sustainability of embedded devices.

We propose the following approaches for energy optimization based on the insights obtained from the energy analysis. The proposed approaches emphasize the need for lightweight security and analysis of energy-performance-security trade-offs for embedded devices.

**Application-aware Obfuscation:**

The primary purpose of energy analysis of different techniques for code obfuscation is to develop an energy profile of different transformations depending on the embedded platform. Currently, open-source tools for code obfuscation such as tigress [8] are not designed for embedded devices. Thus one of the approaches is to port source code from tigress to embedded devices for energy profiling. Integrating energy profiles into tigress would enable developers to develop lightweight transformations depending on the needs of the applications and the overall system-level power budget.

**Obfuscation vs Encryption:**

While obfuscation provides time-limited protection for embedded devices, encryption enhances security through periodic renewal of keys combined with multiple cryptographic algorithms. There are trade-offs associated with obfuscation and encryption for embedded devices and both depend on the size of the source code that needs to be obfuscated or encrypted. Although encryption can be part of obfuscation, we have considered them to be separate in our analysis. Our goal is to assess the energy benefits of obfuscation and encryption and analyze their applicability in embedded applications considering their energy profile and the effectiveness of security. In addition, approaches for integrating obfuscation and encryption in an energy efficient manner can be explored.

**Energy-Performance-Security trade-offs:**

While energy and performance of algorithms for code obfuscation can be estimated, estimating security is of increasing importance towards analyzing the energy-performance-security trade-offs for embedded applications. However, estimating security is challenging due to the lack of available metrics. One of the approaches to estimate security is through approximation [26], where a regression-based model is constructed based on different security parameters pertaining to each of the obfuscation based techniques that are considered independent variables and the resulting security to be the dependent variable. While the approach presented in [26] shows promise, security for different obfuscation techniques need to be estimated and the resulting

### Table 3: Pearson Correlation for Normal and Obfuscated Code

<table>
<thead>
<tr>
<th>App</th>
<th>Lexical</th>
<th>Data</th>
<th>Control</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matrix</td>
<td>0.9977</td>
<td>0.9903</td>
<td>CF:0.9950</td>
<td>0.9925</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DI:0.9946</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>EL:0.9934</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LT:0.9956</td>
<td></td>
</tr>
<tr>
<td>BasicMath</td>
<td>0.9811</td>
<td>0.9528</td>
<td>CF:0.9005</td>
<td>0.8871</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DI:0.9827</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>EL:0.9947</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LT:0.9812</td>
<td></td>
</tr>
<tr>
<td>BitCount</td>
<td>0.9922</td>
<td>0.9445</td>
<td>CF:0.9874</td>
<td>0.9575</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DI:0.9571</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>EL:0.9909</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LT:0.9914</td>
<td></td>
</tr>
<tr>
<td>Vernam</td>
<td>0.9974</td>
<td>0.9866</td>
<td>CF:0.9955</td>
<td>0.9852</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DI:0.9806</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>EL:0.9903</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LT:0.9948</td>
<td></td>
</tr>
</tbody>
</table>

CF: Control Flow Flattening DI: Deadcode Insertion EL: Extended Loop Condition LT: Loop Transformation
energy-performance-security trade-offs analyzed.

7. CONCLUSION

In this work, we analyze the impact of code obfuscation on energy usage for embedded devices. Particularly, we model algorithms for code obfuscation such as Lexical, Data and Control obfuscation and measure the energy and performance of normal and obfuscated code using a measurement tool. Our analysis on a set of benchmarks reveals that while lexical obfuscation has a minimal impact on power consumption, data and control obfuscation was shown to have a significant impact in terms of performance and power consumption. The insights obtained from our analysis can be used for application-aware obfuscation, assessing the energy benefits between obfuscation and encryption and analyzing the energy-performance-security trade-offs.

8. REFERENCES