Peapods:OS-Independent Memory Confidentiality for Cryptographic Engines

LI CONGWU, State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences
LIN JINGQIANG, State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences
CAI QUANWEI, State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences
LUO BO, Department of Electrical Engineering and Computer Science, the University of Kansas

Cryptographic algorithms are widely used in cryptographic systems and play a significant role.The cryptographic software system is linked to the user's core interests. When users use online transactions, e-mail, and remote login services, they rely on the protection of cryptographic software systems. Once the security of cryptographic software systems is not guaranteed, it will cause extreme problems for users. Big losses; on the other hand, the speed of password computing directly affects the user's experience with the crypto software system. Therefore, it is imperative to provide secure, high-speed and reliable cryptographic services.However, existing programs face various memory disclosure attacks and the security of sensitive data cannot be guaranteed, especially key security.In this paper, We present Peapods, a compiler enhancement tool that provides security enhancements for cryptographic software systems that can defend against memory disclosure attacks.A Peapod represents a protected key-related calculation (especially the private key calculation in the public-key cryptographic algorithm).Through the transactional memory mechanism, we can guarantee that in the course of calculation, once an attacker reads the key, the key will be automatically cleared.If the computing key is in a non-calculated state, it is in a ciphertext state.Further, to ensure that calculations can be completed in the transactional memory protection state, it can be split into multiple partitioned tasks.Between the partitioned tasks will exit the transactional memory state, but the calculated intermediate variables will also be encrypted to avoid the leakage of sensitive information.Our experiments show that there is no problem with security and compared to PolarSSL RSA private-key calculation protected by Peapods and non-protected PolarSSL, the performance loss is only 10% within an acceptable range.

CCS Concepts: •**Computer systems organization** →**Embedded systems;** *Redundancy;* Robotics; •**Networks** →Network reliability;

Additional Key Words and Phrases: Implementation of Cryptographic Algorithm, Software Memory Attack, Cold Boot Attack, Transactional Memory, Sensitive information protection

This work is supported by XXX.
 1559-1131/2010/3-ART39 $15.00
DOI: 0000001.0000001

ACM Reference format:
Li Congwu, LIN Jingqiang, CAI Quanwei, and LUO Bo. 2010. Peapods:OS-Independent Memory Confidentiality for Cryptographic Engines. *ACM Trans. Web* 9, 4, Article 39 (March 2010), 18 pages.
DOI: 0000001.0000001

1 INTRODUCTION

Cryptographic algorithms are widely used in cryptographic systems and play a significant role.When we use cryptographic algorithms to implement cryptographic systems, we usually use existing Open Source cryptographic libraries to complete the required calculations and combine them with application functions. The cryptographic software system is linked to the user's core interests. When users use online transactions, e-mail, and remote login services, they rely on the protection of cryptographic software systems. Once the security of cryptographic systems is not guaranteed, it will cause extreme problems for users.On the other hand, the speed of computing directly affects the user's experience with the cryptographic system. Therefore, it is imperative to provide secure, high-speed and reliable cryptographic services. In recent years, many memory protection schemes have been proposed. Copker Mimosa RegRSA . Although these schemes can protect memory information in some scenarios, there are some shortcomings:(1)Difficult to implement under the Windows OS, because users can not get enough operating system permissions, (2) The amount of kernel code increases, and the entire system is vulnerable to attack. Therefore, existing application layer cryptographic software system still face various memory disclosure attacks and the security of sensitive data cannot be guaranteed, especially key security.We need a more principled approach to secure the key in the application layer cryptographic software system.

Peapods Peapods is an automated compiler enhancement tool based on LLVM that provides security enhancements for application layer cryptographic systems that can defend against software memory attacks and cold-boot attack.Peapods protects the key calculation process through Intel TSX.

In our scenario, a Peapod represents a protected key-related calculation (especially the private key calculation in the public-key cryptographic algorithm).Through the transactional memory mechanism, we can guarantee that in the course of calculation, once an attacker reads the key, the key will be automatically cleared.If the computing key is in a non-calculated state, it is in a ciphertext state.Further, to ensure that calculations can be completed in the transactional memory protection state, it can be split into multiple partitioned tasks.Between the partitioned tasks will exit the transactional memory state, but the calculated intermediate variables will also be encrypted to avoid the leakage of sensitive information. Although peapods is mainly dedicated to the protection of keys in cryptographic systems, peapods can also protect user sensitive information when necessary.

Implementation

We implemented Peapods in LLVM for x86 and recompiled PolarSSL and all of its dependencies.We use Peapods to protect Polarssl RSA decryption calculations and users only need to make a few code changes(less than 1000 lines).Experiment shows that compared to PolarSSL protected by Peapods and non-protected PolarSSL, the overhead is only 10% within an acceptable range.

Our contributions

(1) We propose a scheme that can resist memory disclosure attacks in user-mode. (2) Based on LLVM, we implemented Peapods, a tool that automatically implements key protection

and it is independent of the cryptographic algorithm and operating system, programmers do not need to refactor the entire cryptographic software, just add a small amount of code, peapods can provide security enhancements for various cryptographic software systems.The experimental evaluation showed that peapods only introduces a small overhead.

2 BACKGROUND

2.1 Memory Disclosure Attacks

There are many security threats when the cryptographic algorithm is running. One of them is sensitive information stored in the memory, such as a key, which is stolen by an adversary.One of the attacks is a memory disclosure attack. Memory disclosure attacks are read-only memory attacks where the adversary can obtain the contents of the memory but cannot tamper with it. Memory disclosure attacks include software attacks and hardware attacks.Software attack means that the attacker does not physically touch the attack target. It uses software vulnerabilities to bypass the system protection mechanism and illegally accesses the data in the memory.Such as OpenSSL heart bleeding attack, using OpenSSL heartbeat vulnerability, so that remote attackers can obtain a server 64KB memory data by constructing a malicious request.Hardware attack means that an attacker can physically contact an attack target and use a physical method to read the memory of the target.Such as cold-boot attack, using the delay disappearance effect of DRAM, directly read the contents of the memory chip.In short, memory disclosure attacks pose a serious threat to plain-text sensitive data in memory.

physical memory attack

Physical memory attack refers to the behavior of the attacker directly reading the memory chip and obtaining the data illegally after the attacker has physical control over the attack target.In this case, any software-based protection will fail because the attacker can directly manipulate the hardware.Physical memory attack mainly include two types: (1)cold boot attack, which uses DRAM's delay-disappearing effect to directly read the contents of the memory chip, (2)DMA attack, which initiate illegal DMA requests through malicious peripherals, and directly read data from the memory chip.

software memory attack

Software memory attack refers to the behavior of the attacker does not physically contact the attack target, and uses software vulnerabilities to bypass the system protection mechanism and illegally access the data in the memory.For example, OpenSSL heart bleeding attack exploits the vulnerability of OpenSSL heartbeat to enable remote attackers to obtain 64KB of memory data from the server once by constructing a malicious request.Software memory attack mainly include four types: (1)attack based on an isolation mechanism vulnerability, which exploits the defect of the virtual memory subsystem of the operating system to illegally access data from the memory chip, (2)attack based on unclear dynamic memory, which originates from the fact that the dynamic memory storing sensitive information is not cleared and is used by other programs, (3)attack based on cryptographic software's own vulnerability, such as OpenSSL heart bleeding attack, (4)attack based on memory data diffusion, which exploits normal operating system functionality.For example, kernel dumps were originally used to save memory images to disk for developers to debug after a software crash, but attackers can also use this feature to deliberately trigger an exception that causes the software to crash, and then read the sensitive data contained in the process from the disk.

2.2 TSX

Transactional memory is a software technique that simplifies writing concurrent programs.The main idea is to declare a region of code as a transaction. A transaction executes and atomically commits all the results to memory when the transaction succeeds or aborts and cancels all the results if the transaction fails. Transactional memory mechanism provide the Atomicity, Consistency and Isolation qualities. These transactions can safely execute in parallel, and only serial execution is performed in the case of data conflicts, which happen when several threads access the same memory address at the same time and at least one is write operation.When a data conflict occurs, it will cause a rollback and all the results will be canceled.

The recent development is Intels TSX and the implementation in Haswell.Haswell is the first x86 processor to feature hardware transactional memory.Programmers only need to specify critical sections for transactional executions, the processor transparently performs data conflict detection, transaction commit and rollback.TSX tracks the read-set and write-set of a transaction, at a 64B cache line granularity. The read-set and write-set are respectively all the cache lines that the transaction has read from, or written to during execution. A transaction encounters a conflict if a cache line in its read-set is written by another thread, or if a cache line in its write-set is read or written by another thread.

TSX provides two interfaces for programmers to make use of transactional memory. The first mode is Hardware Lock Elision(HLE), provides a pair of compiler hints:xacquire and xrelease to enter or exit the critical section.Once abort happens, the processor will roll back to the original state, and then automatically restarts the execution in a legacy manner. The second mode is Restricted Transactional Memory (RTM), provides three new instructions, XBEGIN, XEND and XABORT, to start, commit, and abort a transactional execution.When using the RTM interface, the programmer needs to specify a fallback handler for the XBEGIN instruction as a parameter and write the fallback handler logic. When a transaction aborts, the program jumps to the specified fallback handler and executes the corresponding program logic.

2.3 LLVM

LLVM is an abbreviation of low level virtual machine. It is actually a compiler framework. With the continuous development of this project, llvm has been unable to fully represent this project, but this name has been continued, and it can now be understood as a full-featured compiler. LLVM can be seen as a collection of compiler and toolchain technologies, and they are modular and reusable.

The traditional compiler is divided into front-end, optimizer, back-end three stages, llvm is also divided into three stages, but slightly different in the design.

Because LLVM is a virtual machine, it has its own Intermediate Representation. It needs to compile the LLVM byte code into a platform-specific assembly language before it can be run by the native assembler and linker to generate assembly code, executable shared libraries, etc.

The front end fetches the source code and then turns it into an intermediate byte code representation.This translation simplifies the work of other parts of the compiler so that they do not need to deal with all the complexity of C++ source code.

In LLVM Optimizer, PASS transforms the program between the intermediate representation.In general, the process is used to optimize the code, that is, the process output program is completely functionally identical to its input program, but the performance is

improved.However, we can also add code logic by adding an IR PASS to achieve a functional improvement over the input program.LLVM provides a lot of APIs for manipulating intermediate bytecode representations, so we can use these interfaces to generate IR directly in memory and run directly to achieve programmatic changes.The intermediate representation of the source code can be further divided into modules, functions, basic blocks, and instruction four-layer structures.

The back-end part translates the intermediate representation into the assembly language of the target platform and generates the actual running machine code.

3 DESIGN

3.1 THREAT MODEL AND SECURITY GOALS

THREAT MODEL:In this paper we assume a powerful attacker who has the ability to read arbitrary areas of memory and overwrite all writable areas of memory.However,as we are concerned with memory disclosure attacks in this paper, the attacker is unable to write to executable memory (marked read-only).Since the registers set to the LLVM compiler reserved registers will be removed from the register allocation pool, the attacker can not read the value of special registers which LLVM compiler reserves either.

Peapods is focused on protecting user-level programs such as PolarSSL RSA decryption calculation, and we do not address protecting the kernel in this paper. We assume that the kernel does not save any user-level registers at least the ones that are used to store the keyduring context switches in user accessible memory. This is true of all major modern operating systems that we are aware of. Custom user-level threading libraries may also require changes to ensure these registers are not saved.

SECURITY GOALS:Under the premise of the above thread model, we proposes Peapods, a LLVM-based compiler enhancement tool that automates the protection of sensitive information in memory to achieve the following security goals:

(1) During key computation, no process other than peapods can access the sensitive data in memory, including the AES master key,the plaintext key and intermediate states.Further, in a transaction execution, whether it is a successful commit or an unexpected termination, each peapods will immediately erase all sensitive data, so the attacker can not access sensitive data by interrupting peapods.At the same time, The sensitive data never appear on the RAM chips.
(2) Without using kernel features, peapods protects the key in every key computation.
(3) Implementation is independent of the cryptographic algorithm. Peapods automatically protects the key computation after the user adds a small amount of code(such as the keyword to tag key and the code to start and end key computation,etc).

The first goal can resist memory disclosure attacks (in kernel-mode), the second goal is to provide support for the application layer cryptographic software systems in user mode, the third goal is to enable Peapods to implement security enhancements for a variety of cryptographic software systems automatically.

In summary, Peapods can resist various memory disclosure attacks and provide security enhancements for various application layer cryptographic software systems, automatically.

3.2 Structure of Peapods

Mimosa provides a defense against memory disclosure attacks in kernel-mode.Based on Mimosa, we propose Peapods to defend against memory disclosure attacks in user-mode.Peapods is a LLVM-based compiler enhancement tool that automatically protects memory that stores

sensitive information, programmers do not need to refactor the entire cryptographic software, just add a small amount of code, peapods can provide security enhancements for various cryptographic software systems.We provide users with a new **keyword** to tag sensitive variables, after the user completes the tag, peapods will automatically protect sensitive variables. During Automated process, we need to protect the following three sensitive variables separately:(1) local variable that address was already determined at compile time, (2) global variable or static variable that address was already determined at compile time, (3) variable that address was not determined at compile time.In order to reduce the impact of time interrupts on transactions, we introduced transaction split.Frequent page faults occur during program execution, we designed efficient preload for Peapods to solve page fault problems.

In order to achieve security goals,we divide the operation process of Peapods into three phases:Programming phase, Compilation phase and Program execution phase.

Programming phase:For sensitive variables whose compile-time address has already been determined, the user needs to add the keywords we provide in the place where the sensitive variables are defined. For sensitive variables whose compile-time address is not determined, the user needs to provide information of such sensitive variables,and then they also need to use the interface we provided for parameter passing through specified structure.When it is necessary to assign a value to a sensitive variable, the user needs to call T_START(args_list) before the sensitive variable assignment and T_END(args_list) after the assignment.

Compilation phase:1.Add initialization logic.2.For sensitive variables whose compile-time address has already been determined, Peapods recognizes variables that have been marked as sensitive data. If it is a local variable, it inserts the XBEGIN instruction before the instruction it defines, and other instructions after the next instruction of its defined instruction, including: (1) Use the master key to perform the CBC mode AES encryption of the intermediate byte code for the variable, and (2) the XEND instruction. Therefore, when these local variables are initialized in the way defined, the sensitive variables are stored in ciphertext in memory. Before the XEND instruction, Peapods will carefully clear the intermediate states used in this stage; If it is a global variable, the AES encryption of the CBC mode of all recognized global variables is completed at the time of program initialization.3.Peapods automatically encrypts and decrypts all sensitive variables in the structure for sensitive variables whose compilation addresses are undetermined.4.START,END, TEM function also adds automatic encryption and decryption protection logic

Program execution phase:We can divide the program execution phase into initialization phase and protection calculation phase.

We use an AES master key to protect user sensitive data. The AES master key is generated at program startup and is then stored in the LLVM compiler reserved register.Sensitive variables are AES encrypted after user-defined initialization or after assignment. When receiving a request for calculating sensitive information, Sensitive information is dynamically generated, used, and destroyed in transaction execution.When there is no computing task, sensitive information is always stored in ciphertext in memory.

Initialization phase: This stage starts when the program starts, mainly performing the following steps: 1.Randomly generate the AES master key and copy it to the reserved register in each CPU core. 2.Call the main function's preload function. 3.Call memset and other library functions. 4.The global variable IV is randomly generated, and the CBC-mode AES encryption is performed on all the global variable-sensitive information that has been identified.Finally, the intermediate states used in this phase are carefully cleared. 5.After

variable initialization and variable assignment,Peapods encrypts the sensitive variable in plain text using the AES master key, so that the sensitive variable is stored in memory in ciphertext form.

The first step is to generate the master key, the second and third steps are to prevent Peapods from abort the transaction due to page fault in the code segment, the fourth step is to store the sensitive information in ciphertext in memory, the last step makes sensitive variables stored in memory in ciphertext after initialization and assignment.

Protection calculation phase: When Peapods receives a sensitive information calculation request, sensitive information is calculated using the corresponding sensitive information, and then the result is provided to the user.This stage contains the following steps: 1.TSX begins to track memory access in the L1 data cache (maintaining read/write sets). 2.Ciphertext sensitive information is loaded into cache from memory. 3.Master key is loaded into cache from reserved register. 4.Use master key to decrypt sensitive information. 5.Calculation. 6.Clear all sensitive information in the cache and registers except the final calculation. 7.Commit transaction

All memory accesses in Protection calculation phase are monitored by hardware. In particular, we use Intel TSX technology to declare a transaction area in which operations that violate the security principles of Peapods are discovered: (1) Any attempt to access changed memory, because the decrypted plaintext Sensitive information and any private intermediate results are in the TSX write set; (2) Data is synchronized from cache to memory due to cache reclaim or replacement

If the above memory exception does not occur, the entire transaction is committed and the results are returned.Otherwise, the hardware's termination processing logic automatically discards all updated memory and then performs a rollback handler to handle the exception.We will immediately retry sensitive information calculations.

4 IMPLEMENTATION

Our system is a C/C++ compiler built on the Clang/LLVM compiler framework. Any application wishing to be protected by Peapods must be recompiled along with all of its dependencies.

The LLVM module pass provides the sensitive variable identification logic, sensitive information encryption and decryption logic we need.

4.1 Master key protection

The master key needs to be saved in the system for a long time.The master key must be tightly protected, otherwise it will pose a threat to data security throughout the system.As persistent sensitive information, the master key is always stored in secure storage in clear text whether or not the sensitive information calculation is to be performed.For physical memory attacks in memory disclosure attacks, any storage area that can restrict access only at the software level is insecure. Secure storage should have physical mechanisms to restrict access.Each CPU core or each hardware thread has a separate set of registers that can only be read by the currently executing instruction.Therefore, the register is the storage resource that runs the program exclusive and it is an ideal secure storage. At the same time, XMM registers are mainly used for floating point and vector operations. Excluding a small number of XMM registers does not have a significant impact on performance in non-massive floating point computing environments.Therefore, we decided to store the master key in the XMM register.

The master key is randomly generated by the rdrand instruction at program startup.Then it is stored in the XMM7 register without any memory operations.In the LLVM backend, we set the XMM7 register as a reserved register.At last,user-level threading libraries have also made corresponding changes to prevent the value of XMM7 register from leaking into memory.

4.2 sensitive variables Identification

We divided the sensitive variables into three categories: (1) local variable that address was already determined at compile time, (2) global variable or static variable that address was already determined at compile time, (3) variable that address was not determined at compile time. For sensitive variables whose compile-time addresses have been determined, such as sensitive variables for array types, structure types (members without pointer-type variables), we provide the user with the new **keyword**: attribute ((annotation "Private Key")) when the user defining a sensitive data variable of this type, they need to add the keyword to the variable definition.Peapods will identify the variable that has been marked as sensitive data. If it is a local variable, it inserts the XBEGIN instruction before the instruction it defines, and inserts the instruction following its defined instruction: (1)The intermediate byte code of the logic of use the master key pair to encrypt this variable with AES encryption of the CBC mode, and (2) the XEND instruction.Therefore, when these local variables are initialized in the way defined, the sensitive variables are stored in ciphertext in memory. Before the XEND instruction, Peapods will carefully clear the intermediate states used in this stage;If it is a global variable or static variable, the AES encryption of the CBC mode of all recognized global variables is completed at the time of program initialization.

For sensitive variables with undecided compile-time addresses, such as pointer-type sensitive variables, since the address and length of sensitive variables cannot be obtained at compile time, the user is required to provide information of such sensitive variables. Then, the user uses the specified structure passes the parameters in the utility function provided by us, and Peapods will automatically encrypt and decrypt all the sensitive variables in the structure.

```
1 struct args{
2     unsigned char *data;
3     int len;
4     struct args* next;
5 }*args_list,args_list_next;
```

4.3 Sensitive information protection

We need to consider time interruptions during the calculation. The calculation of sensitive information (such as private keys) is usually relatively time-consuming. Therefore, the execution of a transaction that would otherwise be successful may be terminated by a time interruption.In addition, other interrupts can also cause the transaction to terminate.The solution to other kernel-mode memoryless encryption schemes is to disable interrupts,such as TRESOR, PRIME, Copker and Mimosa. We propose Sensitive Calculation Split, a method for breaking time-consuming large Calculation into multiple small transactions.

The design of the Sensitive Calculation Split is mainly to achieve the following goals: 1.Even if the entire time-consuming calculation is not completed, we can still save some

time-consuming intermediate calculation results. When the transaction occurs abort, we can use these already calculated intermediate calculation results to start the next calculation. 2.Compared to transactions that were not split prior to the entire calculation, only a small number of CPU clock cycles were consumed in each transaction and only a small amount of memory space was occupied. Therefore, these small transactions are easier to submit successfully.

Since the introduction of transaction splitting, the entire calculation is no longer an atomic operation, we need to encrypt the calculated intermediate calculation results using the AES master key before the transaction ends, and then perform AES decryption after the start of the next transaction to ensure security outside the transaction.These security logic are automatically implemented when the TEM function is called

After we split a PolarSSL RSA decryption calculation into 128 times, the program performs well.

We encapsulate the xbegin and xend instructions in user mode into utility functions T_START and T_END.They are used to start transactions and end transactions respectively.In addition,We also encapsulate an Utility functions T_SPLIT to split sensitive calculation.We track all function call instructions in IR PASS, recognize the call to T_START, T_END, and T_SPLIT. Then we insert the corresponding intermediate representation code of encryption and decryption before the T_END function call instruction, after the T_START function call instruction, and both before and after the T_SPLIT function call instruction respectively.

When users want to assign values to sensitive variables, they need to call T_START(args_list) before assigning sensitive variables and T_END(args_list) after assigning values.In IR PASS, we add logic to automatically encrypt and decrypt sensitive variables in utility functions such as T_START and T_END. In this way, the entire assignment process is performed under transaction protection. After the transaction is started, the sensitive variables are decrypted, and then encrypt sensitive variables before ending the transaction. Therefore, there is always only a ciphertext sensitive variable in memory.

Therefore, for the sensitive variables marked by the user with the keywords we provide, they are stored in the ciphertext in the memory after variable initialization, decrypted into plaintext before the sensitive calculation and assignment, and encrypted into ciphertext after the sensitive calculation and assignment are completed.

It is worth noting that reading and writing files within a transaction will cause the transaction to abort. Therefore, Peapods cannot provide security protection during the assignment phase if the user reads a file to assign sensitive variables.In fact, in this case, the user can still call T_START(args list) after reading the file, call T_END(args list) after assignment, and clear the plaintext intermediate variable to make the sensitive variable exist in ciphertext after the assignment.However, in the process of reading a file, sensitive information is exposed in memory in the form of clear text. It is assumed that there is no memory disclosure attack at this time.At the same time, there is a backup of sensitive information in the hard disk. An attacker can also read the hard disk to obtain sensitive information. The user needs to delete the file containing sensitive information after reading the file.

4.4 Page fault handling

We protected Polarssl RSA decryption calculations with peapods, during the execution of the program, we used perf, a performance analysis tool to help us knowing the program's

Table 1. Utility functions list

Function Name	Function	Usage Method	Implementation logic
T_START(args_list)	Start a transaction	Used before sensitive operations begin	(1)XBEGIN(2)get the master key from the XMM7 register(3)generate AES context using the master key(4)AES decryption of CBC mode for identified sensitive data variables
T_END(args_list)	End a transaction	Used after sensitive operations end	(1)AES encryption of CBC mode for identified sensitive data variables(2)Erase intermediate results(3)XEND
T_SPLIT(args_list)	Split a transaction	Used in locations where transactions need to be split within the transaction	(1)AES encryption of CBC mode for identified sensitive data variables(2)Generate intermediate variable IV, then randomly initialize and perform AES encryption of CBC mode for the intermediate variables passed in by the user.(3)Erase intermediate result-s(4)XEND(5)XBEGIN(6)get the master key from the XMM7 register(7)generate AES context using the master key(8)AES decryption of CBC mode for identified sensitive data variables

operating status. Perf relies on the Intel performance monitoring facility, which supports precision-event-based sampling. This function can record the current processor state when a specific event occurs.We monitored the RTM_RETIRED.ABORTED event, which was triggered when RTM execution was terminated.Based on the processor status acquired, we can find out the cause of the termination.Perf reported a lot of the termination of the transaction caused by the pages fault. After analysis and exploration, we found that these pages fault mainly come from pages fault of the code segment.

For the pages fault of the code segment, Peapods implements automatic preloading of the code to be executed within the transaction based on LLVM.The basic principle is to insert an empty function before and after the function definition of the function to be executed in the transaction, and to call all the inserted empty functions in the program initialization phase. At this point, the code segment of the function to be executed in the transaction is loaded into memory along with the code segment of the empty function.We traverse all functions in IR PASS and add the corresponding before and after function definitions before and after the function definition.Then, tracking the function calls in all functions, in the corresponding before and after functions, adding the before and after function calls of the functions called in the original function.When there is no function call in the function, only two empty functions are defined.Finally, we add calls to main_before and main_after during

the initialization phase of the program to implement the call to all before and after functions during the initialization phase.

The above implementation still has some problems:

1.Due to the need to add a new function definition at the location of the function definition, the current solution cannot support automatic preloading of the library function. Our solution is to add a call to a library function such as memset when the program is initialized.For the same reason, the current solution cannot support the automatic preloading of LLVM functions such as llvm.var.annotation. Calling these functions will not result in page fault. We have solved this problem by adding a whitelist filter solution.

2.The current scenario cannot support automatic preloading of functions whose code segment size exceeds one page.According to the pre-loading mechanism, only the contents of the same page as the before and after functions are loaded into memory. If the size of the code segment of a single function exceeds one page, some contents may still not be loaded into memory. This situation will still result in pages fault within the transaction.

4.5 usage method

Users only need to add a small amount of code when using peapods to automatically implement sensitive information protection.When a user needs to add security enhancements to a certain cryptographic software system, (1) the __attribute__(annotation "Private_Key") keyword needs to be added before the definition of the sensitive variable of the type whose address has been determined during the compilation, (2) assign pointer-type sensitive variables to the args_list structure, then pass parameters to the START(args_list) and END(args_list) functions, (3) The START(args_list) function needs to be called before the variable assignment, and the END(args_list) function is called after the assignment.

In addition, the user needs to call the START(args_list) function before calculating the sensitive information and call the END(args_list) function after the sensitive information is calculated.When the sensitive information calculation is too time-consuming, the user needs to call the TEM(args list) function at the split location to split the transaction.

After users complete the above work, Peapods will automatically use HTM to protect sensitive information. Therefore, we can provide security enhancements for various types of cryptographic software systems without refactoring the entire cryptographic software in the presence of security vulnerabilities.The amount of code that the user needs to add is only a few hundred lines to several thousand lines (depending on the number and type of sensitive variables, the number of splits, etc.), the proportion compared to the total code volume of the cryptographic software system (usually several hundred thousand lines) is small.

It is worth noting that all user programs in the system need to be compiled with our modified LLVM compiler to ensure the security of the master key.

```
1  __attribute__((annotation"Private_Key")) unsighed char secret[16];
2
3  T\_START(args_list);
4      secret={\
5          0x45,0x46,0x88,0x32,\
6          0x2a,0x6d,0x8c,0x31,\
7          0x58,0xf2,0x30,0x02,\
8          0x4f,0x32,0x7d,0x22\
9      };
10 T\_END(args_list);
11
```

```
12  T\_START(args_list);
13      secret_protect_compute_1();
14  T\_SPLIT(args_list);
15      secret_protect_compute_2();
16
17      ...
18
19  T\_SPLIT(args_list);
20      secret_protect_compute_n();
21  T\_END(args_list);
```

5 SECURITY DISCUSSION

First, peapods have their own limitations.One limitation of Peapods is that it requires all user-mode executives must be recompiled by the modified LLVM we provide.This is conducive to the formation of a safe execution environment.It is difficult for an attacker to execute his own attack code, because it requires breaking the control flow integrity first.There are many effective control flow integrity protection implementation proposed, such as Cryptographic CFI, Forward Edge CFI, etc.Another limitation is that Peapods cannot protect libraries without source code.This problem could be solved when library developers apply Peapods to their closed-source libraries.

During the sensitive information calculation phase, Peapods relies on Intel TSX to ensure the confidentiality of protected sensitive information. However, cache(the foundation of TSX implementation) is limited by space and shared by all cores. This might lead to denial of service(DoS) attacks.Most of Intel's CPUs implement 8-way group-associated L1D caches, so 9 memory accesses to the same set of caches will inevitably result in transaction termination.Moreover, Intel does not guarantee that all L1D caches can be used in TSX.In addition, a memory-intensive process may also make it difficult for Peapods to complete a transaction.Due to the shared L3 cache, it is likely to evict the cache line that Peapods is using.Users can use T-SPLIT function to split the sensitive information calculation process, reducing the execution time and memory usage of a single transaction, which can effectively alleviate the impact of DoS attacks on Peapods.We completed the experiment to verify this result.

When the computer undergoes a core dump, as the Peapods transaction is interrupted, the master key, plaintext sensitive information, and intermediate calculations that have already been calculated are all cleared.Therefore, the attacker still cannot obtain sensitive information in plaintext.

Attackers may also attack sensitive information through the side channel.Fortunately, Peapods is immune to cache-based time-side channel attacks because AES-NI itself is not subject to any known side-channel attacks and the sensitive information calculation is fully implemented in the cache.Peepods can't resist timing side channel attacks, but programmers can defend against the attacks by refactoring the code and adding blinding algorithm.

6 EVALUATION

This section presents the experimental results by measuring the performance of Peapods. The experimental machine has an Intel Core i7-4770S quad-core processor running a Linux operating system (kernel version 3.13.1).We experimented with the PolarSSL cryptographic algorithm library and used Peapods to protect 2048-bit RSA private key calculations after splitting into 128 transactions and 256 transactions.The comparison objects in the experiment

include: (1) the official default configuration of PolarSSL (version 1.3.9), (2) Peapods_128, the Peapods split into 128 transactions,(3) Peapods_256, the Peapods split into 256 transactions.

6.1 Local Performance

The local throughput for PolarSSL, Peapods_128, and Peapods_256 are: 391/s, 341/s, and 352/s (4 threads).Compared to PolarSSL, the performance of Peapods_128 and Peapods_256 dropped by 12.8% and 10.0%, respectively.The performance overhead introduced by Peapods mainly comes from: (1) waste of CPU clock cycles due to rollback of transactions; (2) encryption and decryption protection of sensitive information and intermediate calculation results; (3) preloading.The reason Peapods_256 performs slightly better than Peapods_128 is that the more granular splitting reduces transaction abort due to time-outs and other factors.

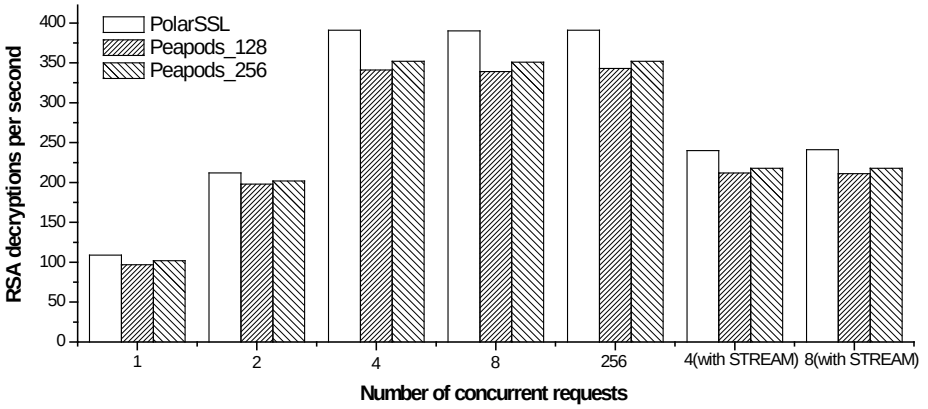

Fig. 1. Local Performance

We also measured whether a memory-intensive program will have a greater impact on Peapods by running the Geekbench 3 memory stress test while Peapods is performing RSA private key calculations.In the experiment, 4 different memory stress tests will run on all CPU cores, which will cause about 10GB/s of memory data transfer.The maximum memory transfer rate supported by the machine when no user program is running is 13.7GB/s.Peapods_128 dropped from 341 RSA decryptions per second to 212 RSA decryptions per second, performance decreased by 37.8%, Peapods_256 decreased from 352 RSA decryptions per second to 218 RSA decryptions per second, performance decreased by 38.1%.PolarSSL decreased from 391 RSA decryptions per second to 241 RSA decryptions per second, performance decreased by 38.4%.

In short, the performance overhead of Peapods is acceptable. At the same time, Peapods perform better than other transaction-based protection due to preloaded optimizations.

6.2 Factors affecting transaction abort

We also evaluated the impact of fine-grained sensitive computational process splitting on transaction abort. We use peapods to protect a simple program that is set to a transaction area after processing, then constantly adjust the memory usage and transaction time in the transaction through write operations and loop instructions. We measured the duration of each transaction by instrumenting the source code that begins and ends transactions

with rdtsc instructions.Table 2 and Table 3 display the effect of transaction time and memory usage on transaction abort.For each transaction, we conducted 200 experiments, the data(abort rate) in the table is the average of the experimental results.

Table 2. Effect of transaction time and memory usage on transaction abort(single thread)

Time(cycles) / **Memory(bytes)**	3000	30000	120000	210000	300000
0	0	1.14%	3.92%	6.94%	8.23%
300	0.54%	1.35%	4.42%	7.36%	8.85%
1200	0.64%	1.72%	4.69%	7.26%	9.72%
2100	0.68%	2.03%	4.91%	8.18%	10.25%
3000	0.68%	2.18%	5.23%	8.42%	11.45%

Table 3. Effect of transaction time and memory usage on transaction abort(4 threads)

Time(cycles) / **Memory(bytes)**	3000	30000	120000	210000	300000
0	0	1.62%	5.12%	8.96%	12.23%
300	0.76%	2.17%	6.37%	9.43%	12.34%
1200	1.14%	2.93%	7.67%	10.3%	14.63%
2100	0.98%	2.93%	8.46%	13.64%	17.35%
3000	1.28%	3.88%	9.76%	14.92%	18.05%

It is not difficult to find that transaction time and memory usage have an impact on the transaction abort.The impact of transaction time on transaction abort is very intuitive. As the transaction time increases, the transaction abort rate increases greatly.When the memory usage is small, the change in memory usage has a greater impact on the transaction abort. As the memory usage increases, this effect is gradually reduced.When the memory usage increases to a certain value, the transaction abort rate increases rapidly (not shown in the tables), which is caused by the cache capacity limit.

6.3 Impact on Concurrent Processes

We used Geekbench 3 to test Peapods' impact on other concurrent processes.Figure X shows Geekbench 3 scores in multi-core mode and single-core mode.The baseline is a score in clean environment.

The Geekbench 3 benchmark program performed integer, floating point, and memory bandwidth tests, respectively.PolarSSL, Peapods_128, and Peapods_256 experiments have similar scores. Unprotected PolarSSL is slightly better than Peapods.When Geekbench occupies multiple cores, the load change that simultaneously processes RSA private key calculation requests cannot be ignoredthe baseline has a clear demarcation from other scores.It is worth noting that XMM registers are mainly used for floating-point calculations, and generally only 1-2 XMM registers are used.When the environment has only a small number of floating point calculations, we will not significantly affect the XMM7 register from the register allocation pool. When there are a large number of floating point calculations in the environment, the overall performance will be slightly reduced.In short, Peapods do not have a significant additional impact on other concurrent processes.

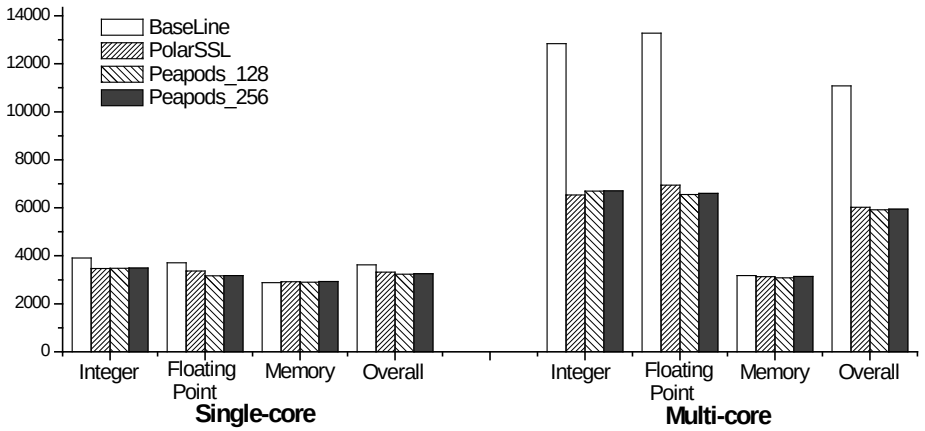

Fig. 2. Impact on Concurrent Processes

6.4 SGX

7 RELATED WORK

In this section, we discuss a number of important studies that are related to Peapods.

Attack and Defense on Cryptographic Keys. Memory disclosure attacks pose a serious threat to cryptographic keys.Harrison et al. provide ways to keep only one copy of cryptographic keys in allocated memory[17].Scrash removes sensitive data from crash reports in the case of program failures to avoid sensitive data being leaked to disks[4].Shreds[6] is proposed to limit the memory that each thread can access.It is a general software key protection implementation. Henson et al. use special hardware iRAM to implement full-disk encryption[18].

Müller et al. propose TRESOR[11] to prevent the cold-boot attacks on full-disk encryption, by storing AES keys in registers only.Using the AES key protected by TRESOR as a key-encryption key, PRIME[12] and RegRSA[41] implemented the RSA computations in registers,RegRSA uses vector instructions in calculations, it provides much better performance than PRIME. Copker[13] and Mimosa[14] implemented the RSA computations in caches.Mimosa can defend against both software memory attacks and cold-boot attacks due to the use of Intel TSX[20].Ramcrypt[10] is a memory encryption implementation. When the program is running, only the page (4KB) that the program is accessing is decrypted. Mashtizadeh et al. propose CCFI[28] to CFI based on cryptographic message authentication codes and the implementation stores AES keys in user-mode registers.

Intel Software Guard eXtensions (SGX) provide a hardware-enabled secure container that is isolated from other processes[21]. Confidentiality and integrity of the protected process will be maintained even in the presence of privileged malware. SGX shows the same tendency and potential as TSX that secure systems can be built on top of hardware features.

Transaction Memory Implementation. Intel TSX features support in the fourth generation of Core CPU (Haswell architecture), implementing hardware transaction memory mechanism. In addition to the Intel TSX, there are a variety of hardware implementation transaction memory mechanism design[15][1], also included in SUN SPARC[9], IBM System z[22], IBM Blue Gene / Q[39] and other platforms implementations.The transaction

memory mechanism can also be implemented by software[5][16][30][35][34][19], or a mixed implementation of software and hardware[29][8][25][31]: some functions are implemented by CPU hardware, and some functions are implemented by system software.For distributed computing environments, there is also a corresponding distributed transaction memory mechanism[3][7][24][32][33] for memory read and write control between multiple computers.

Transaction Memory Application. The transaction memory mechanism directly relates to the read and write control of the memory data. At the same time, a large amount of system security is related to the access of the memory data[38].In 2008, TMI[2] completed the implementation of the authorization strategy based on the software transaction memory mechanism.TSX-CFI implements coarse-grained control flow integrity and fine-grained control flow integrity based on RTM and HLE, respectively.TxIntro[27] and Mimosa[14] implement memory data security with memory control of transactional memory mechanisms:TxIntro monitors the abnormal changes in the Read-Set detection data in the transaction memory task, and Mimosa monitors the Write-Set to detect unauthorized read operations in the transaction memory task.HAFT[26] implemented the instruction-level redundancy Instruction-Level Redundancy (ILR) fault-tolerant system by using the transaction memory mechanism's rollback handling.DrK attack[23] exploits the application layer isolation feature of the exception handling of the transaction memory mechanism to attack the Kernel Address Space Layout Randomization (KASLR):even if the user-mode application has an arbitrary number of memory access exceptions, it will not fall into the exception handling of the operating system.Using the above features, T-SGX[36] implements a defense mechanism against malicious operating systems and eliminates the malicious operating system's Page-fault side channel attack on SGX Enclave[40][37].

8 CONCLUSION

The cryptographic systems are now widely used and linked to the user's core interests.However,existing memory protections are difficult to support key security in user-mode cryptographic systems.This paper introduces Peapods, which is a automated, efficient, and OS-Independent compiler enhancement tool that provides security enhancements for application layer cryptographic systems that can defend against software memory attacks and cold-boot attack.It ensures that the key is stored in ciphertext in memory when it is not being calculated and in the course of calculation, once an attacker reads the key, the key will be automatically cleared.Although peapods is mainly dedicated to the protection of keys in cryptographic systems, peapods can also protect user sensitive information when necessary.

ACKNOWLEDGMENTS

The authors would like to thank XXX

REFERENCES

[1] C. Scott Ananian, Krste Asanovic, Bradley C. Kuszmaul, Charles E. Leiserson, and Sean Lie. 2006. Unbounded Transactional Memory. *IEEE Micro* 26, 1 (2006), 59–69.

[2] Arnar Birgisson, Mohan Dhawan, Vinod Ganapathy, and Liviu Iftode. 2008. Enforcing authorization policies using transactional memory introspection. In *ACM Conference on Computer and Communications Security, CCS 2008, Alexandria, Virginia, Usa, October*. 223–234.

[3] Robert L. Bocchino, Vikram S. Adve, and Bradford L. Chamberlain. 2008. Software transactional memory for large scale clusters. In *ACM Sigplan Symposium on Principles and Practice of Parallel Programming*. 247–258.

[4] Pete Broadwell, Matt Harren, and Naveen Sastry. 2003. Scrash: a system for generating secure crash information. In *Conference on Usenix Security Symposium*. 19–19.

[5] Brian D Carlstrom, Austen Mcdonald, Hassan Chafi, Jae Woong Chung, Chi Cao Minh, Christos Kozyrakis, and Kunle Olukotun. 2006. The Atomos transactional programming language. *Acm Sigplan Notices* 41, 6 (2006), 1–13.
[6] Yaohui Chen, Sebassujeen Reymondjohnson, Zhichuang Sun, and Long Lu. 2016. Shreds: Fine-Grained Execution Units with Private Memory. In *Security and Privacy*. 56–71.
[7] Maria Couceiro, Paolo Romano, Nuno Carvalho, and Lus Rodrigues. 2009. D2STM: Dependable Distributed Software Transactional Memory. In *IEEE Pacific Rim International Symposium on Dependable Computing*. 307–313.
[8] Peter Damron, Alexandra Fedorova, Yossi Lev, Victor Luchangco, Mark Moir, and Daniel Nussbaum. 2006. Hybrid transactional memory. In *International Conference on Architectural Support for Programming Languages & Operating Systems*. 336–346.
[9] Dave Dice, Yossi Lev, Mark Moir, and Daniel Nussbaum. 2009. Early experience with a commercial hardware transactional memory implementation. *ACM SIGPLAN Notices* 44, 3 (2009), 157–168.
[10] Gabor Drescher and Michael Backes. 2016. RamCrypt: Kernel-based Address Space Encryption for User-mode Processes. In *ACM on Asia Conference on Computer and Communications Security*. 919–924.
[11] Felix C. Freiling and Andreas Dewald. 2011. TRESOR runs encryption securely outside RAM. In *Usenix Conference on Security*. 17–17.
[12] Behrad Garmany and Tilo Mller. 2013. *PRIME: private RSA infrastructure for memory-less encryption*. ACM. 149–158 pages.
[13] Le Guan, Jingqiang Lin, Bo Luo, and Jiwu Jing. 2014. Copker: Computing with Private Keys without RAM. In *Network and Distributed System Security Symposium*.
[14] Le Guan, Jingqiang Lin, Bo Luo, Jiwu Jing, and Jing Wang. 2015. Protecting Private Keys against Memory Disclosure Attacks Using Hardware Transactional Memory. In *IEEE Symposium on Security and Privacy*. 3–19.
[15] Lance Hammond, Vicky Wong, Mike Chen, Brian D Carlstrom, John D Davis, Ben Hertzberg, Manohar K Prabhu, Honggo Wijaya, Christos Kozyrakis, and Kunle Olukotun. 2004. Transactional Memory Coherence and Consistency. In *International Symposium on Computer Architecture, 2004. Proceedings*. 102–113.
[16] Tim Harris, Simon Marlow, Simon Peyton-Jones, and Maurice Herlihy. 2005. Composable memory transactions. 48–60.
[17] Keith Harrison and Shouhuai Xu. 2007. Protecting Cryptographic Keys from Memory Disclosure Attacks. In *Ieee/ifip International Conference on Dependable Systems and Networks*. 137–143.
[18] Michael Henson and Stephen Taylor. 2013. Beyond full disk encryption: protection on security-enhanced commodity processors. In *International Conference on Applied Cryptography and Network Security*. 307–321.
[19] Maurice Herlihy, Victor Luchangco, and Mark Moir. 2006. A flexible framework for implementing software transactional memory. 253–262.
[20] Intel. 2012. Chapter 8: Intel transactional memory synchronization extensions. In *Intel Architecture Instruction Set Extensions Programming Reference*.
[21] Intel. 2014. Intel software guard extensions programming reference. (2014).
[22] Christian Jacobi, Timothy Slegel, and Greiner Dan. 2012. Transactional Memory Architecture and Implementation for IBM System Z. In *Ieee/acm International Symposium on Microarchitecture*. 25–36.
[23] Yeongjin Jang, Sangho Lee, and Taesoo Kim. 2016. Breaking Kernel Address Space Layout Randomization with Intel TSX. In *ACM Sigsac Conference on Computer and Communications Security*. 380–392.
[24] Christos Kotselidis, Mohammad Ansari, Kim Jarvis, Mikel Lujn, Chris Kirkham, and Ian Watson. 2008. DiSTM: A Software Transactional Memory Framework for Clusters. In *International Conference on Parallel Processing*. 51–58.
[25] Sanjeev Kumar, Michael Chu, Christopher J. Hughes, Partha Kundu, and Anthony Nguyen. 2006. Hybrid transactional memory. 209–220.
[26] Dmitrii Kuvaiskii, Rasha Faqeh, Pramod Bhatotia, Pascal Felber, and Christof Fetzer. 2016. HAFT: hardware-assisted fault tolerance. In *Eleventh European Conference on Computer Systems*. 25.
[27] Yutao Liu, Yubin Xia, Haibing Guan, Binyu Zang, and Haibo Chen. 2014. Concurrent and consistent virtual machine introspection with hardware transactional memory. In *IEEE International Symposium on High PERFORMANCE Computer Architecture*. 416–427.

[28] Ali Jose Mashtizadeh, Andrea Bittau, and Boneh Dan. 2015. CCFI:Cryptographically Enforced Control Flow Integrity. In *ACM Sigsac Conference on Computer and Communications Security*. 941–951.
[29] K. E. Moore, J. Bobba, M. J. Moravan, and M. D. Hill. 2006. LogTM: log-based transactional memory. In *The Twelfth International Symposium on High-Performance Computer Architecture*. 254–265.
[30] Yang Ni, Adam Welc, Ali Reza Adl-Tabatabai, Moshe Bach, Sion Berkowits, James Cownie, Robert Geva, Sergey Kozhukow, Ravi Narayanaswamy, and Jeffrey Olivier. 2008. Design and implementation of transactional constructs for C/C++. *Acm Sigplan Notices* 43, 10 (2008), 195–212.
[31] Ravi Rajwar, Maurice Herlihy, and Konrad Lai. 2005. Virtualizing Transactional Memory. 494–505.
[32] Paolo Romano, Luis Rodrigues, and Nuno Carvalho. 2010. Cloud-TM:harnessing the cloud with distributed transactional memories. *Acm Sigops Operating Systems Review* 44, 2 (2010), 1–6.
[33] Mohamed Saad and Binoy Ravindran. 2011. Snake: Control Flow Distributed Software Transactional Memory. In *International Conference on Stabilization, Safety, and Security of Distributed Systems*. 238–252.
[34] Martin Schindewolf, Albert Cohen, Wolfgang Karl, Andrea Marongiu, and Luca Benini. 2009. Towards Transactional Memory Support for GCC.
[35] Nir Shavit and Dan Touitou. 1995. Software transactional memory. (1995), 204–213.
[36] Ming Wei Shih, Sangho Lee, Taesoo Kim, and Marcus Peinado. 2017. T-SGX: Eradicating Controlled-Channel Attacks Against Enclave Programs. In *Network and Distributed System Security Symposium*.
[37] Shweta Shinde, Leong Chua Zheng, Viswesh Narayanan, and Prateek Saxena. 2015. Preventing Your Faults From Telling Your Secrets: Defenses Against Pigeonhole Attacks. *Computer Science* (2015).
[38] Laszlo Szekeres, Mathias Payer, Tao Wei, and Dawn Song. 2013. Eternal War in Memory. In *Security and Privacy*. 48–62.
[39] Amy Wang, Matthew Gaudet, Peng Wu, Jos Nelson Amaral, Martin Ohmacht, Christopher Barton, Raul Silvera, and Maged Michael. 2012. Evaluation of Blue Gene/Q hardware support for transactional memories. (2012), 127–136.
[40] Yuanzhong Xu, Weidong Cui, and Marcus Peinado. 2015. Controlled-Channel Attacks: Deterministic Side Channels for Untrusted Operating Systems. In *Security and Privacy*. 640–656.
[41] Yuan Zhao, Jingqiang Lin, Wuqiong Pan, Cong Xue, Fangyu Zheng, and Ziqiang Ma. 2016. *RegRSA: Using Registers as Buffers to Resist Memory Disclosure Attacks*. Springer International Publishing. 293–307 pages.

Received February 2007; revised March 2009; accepted June 2009