Optimal Parameter Selection for Efficient Memory Integrity Verification Using Merkle Hash Trees

Dan Williams and Emin Gün Sirer
Department of Computer Science
Cornell University
Ithaca, NY 14853
{djwill, egs}@cs.cornell.edu

Abstract

A secure, tamperproof execution environment is critical for trustworthy network computing. Newly emerging hardware, such as those developed as part of the TCPA and Palladium initiatives, enables operating systems to implement such an environment through Merkle hash trees. We examine the selection of optimal parameters, namely blocksize and tree depth, for Merkle hash trees based on the size of the memory region to be protected and the number of memory updates between updates of the hash tree. We analytically derive an expression for the cost of updating the hash tree, show that there is an optimal blocksize for the leaves of a Merkle tree for a given filesize and update interval that minimizes the cost of update operations, and describe a general method by which the parameters of such a tree can be determined optimally.

1. Introduction

Trustworthy network computing fundamentally requires the ability to reason about the state of a computation on remote nodes. Such reasoning relies on two mechanisms. First, a node needs to be able to represent the state of a local computation such that other nodes in the network can make an intelligent decision on whether or not to trust the results of that computation. Previous work on attestation [8, 10, 1] addresses precisely this issue; a certificate chain rooted in secure hardware can attest that a given version of the operating system executed a particular version of an application. Using such a certificate chain, a remote game server, for instance, may decide to permit (or reject) a client attempting to connect to the game with a good (or hacked) game client. Similar intelligent trust decisions on whether a client will obey a desired protocol may be made in other distributed computing settings, including peer-to-peer systems and ad hoc networks, using the same mechanism. In essence, the certificate chain can establish that certain predicates over the code, usually represented compactly through code version numbers or cryptographic hashes, hold at a certain point in time.

But attesting to the state of a client at a given point in time is not sufficient to establish trust. A second mechanism, namely, an isolated, secure, tamperproof execution environment, is required to reason about the state of the computation subsequent to the attestation. In the example above, a game server should allow clients to connect only if their binary is verified to not be hacked at the time of connection (achieved through attestation), and if the connected game client can execute in a tamperproof environment where the binary cannot be modified after connection (achieved through tamperproof execution). This latter mechanism has been the subject of much recent work [12, 11, 18, 6, 7], buoyed by the emergence of hardware support for secure execution in general-purpose computers [19] and industry support for secure execution as in Microsoft's Palladium [4]. All mechanisms for tamperproof execution proposed to date rely on costly cryptographic hashes to detect modifications to memory; however, none minimize the cost of hashing.

This paper focuses on the use of cryptographic hashes to secure memory against unauthorized modifications, and derives an expression for optimal hash parameters. A well known method to ensure that the contents of a data structure stored in untrusted storage (memory, disk or tertiary storage) have not been tampered with is to compute a hash of that data upon creation and store the hash in a secure location. The next time an element in the data structure is used, the hash is recomputed and checked against the stored hash; unauthorized modifications to the data structure will be caught through a hash mismatch. However, this naïve use of hashing can become extremely expensive when used on large data structures.

Proceedings of the Third IEEE International Symposium on Network Computing and Applications (NCA'04)
0-7695-2242-4/04 $ 20.00 IEEE

Merkle hash trees have been proposed as a means to reduce the cost for hashing large data structures[14, 15]. They are used to take a secure summary snapshot of a memory region, which can then be used to detect tampering. A memory region is divided into smaller blocks, the hashes of which form the *leaf hashes* at the leaves of a complete binary tree. The value of an inner node of the tree, an *inner hash*, is obtained by concatenating and hashing the values of its child nodes. After a set of updates to a memory region that constitute a transaction, a new secure summary snapshot of the data structure is obtained by incrementally recomputing the leaf hashes corresponding to the modified blocks, as well as the inner hashes from each modified leaf to the root of the tree. Once a new Merkle hash tree is computed, the hashes can be stored in a secure location, such as a secure coprocessor, and used to ascertain the integrity of the data structure kept in ordinary memory. Overall, Merkle hash trees constitute a very simple and effective way to take a secure summary snapshot of a data structure.

The blocksize is the critical parameter of a Merkle hash tree. A large blocksize reduces the depth of the tree at the cost of increasing the leaf hash cost. A small blocksize makes leaf hashes cheaper to compute, though it also increases the depth of the tree, and correspondingly, the time spent computing inner hashes.

This paper examines the optimal selection of blocksize for Merkle hash trees. We derive an analytical model that describes the cost of incremental updates to a Merkle hash tree given the total size of a memory region to be protected and the number of modified memory locations in each transaction, and we can numerically determine the blocksize that minimizes the cost of performing updates to the tree. This, in turn, enables an efficient mechanism for implementing tamperproof execution using commodity memory and storage devices.

This paper makes two contributions. First, it shows that there is a minimum update cost that can be achieved by a hash tree through careful selection of the blocksize at the leaves of the tree. Second, it derives this optimal blocksize given simple parameters, easily determined in practice. The choice of optimal parameters for tamperproof memory in turn leads to efficient systems for secure, trustworthy execution. Surprisingly, the optimal parameters in many common settings differ from natural choices that designers may be tempted to pick, such as the native cacheline or page size.

In the next section, we discuss related work in the areas of tamper-proof memory and Merkle trees. Section 3 describes our system model to help put the problem in context. An analytical model of the problem results, and implications for implementing tamperproof execution hardware, the cornerstone of trusted network computing, are presented in Section 4, and Section 5 summarizes the contribution and concludes.

2. Related Work

Merkle trees were originally presented as a method in which two entities can agree on a shared secret using a public key infrastructure [14, 15], but have since been used in a variety of other applications including fast digital signature schemes for flows and multicasts [22] and verification of signatures on read only file systems [5] Blum et al. [2] use Merkle hash trees to provide general memory integrity, in a manner similar to the system model used in this paper; however, this work does not examine how to determine the hash blocksize. There has also been some work focused on the integrity of persistent storage in databases [13] and DRM systems [16].

Recent work on trustworthy execution platforms [12, 11, 17, 6, 18, 4] has examined practical mechanisms for attestation [10, 23, 20] and tamperproof execution. This work spans a large space including the design of secure coprocessors and security enhancements to ordinary processors to provide a trustworthy execution environment, the attestation of the underlying system to the integrity of its applications, the structure of the underlying operating system to provide secure attestation, and finally, on the trustworthiness of applications.

The eXecute Only Memory (XOM) architecture [12] provides a trusted environment for applications through additional hardware in the processor that creates an isolated, secure, tamperproof execution environment to applications. The additional hardware encrypts memory and register values as they are transferred into and out of the processor. This additional hardware enables tamperproof execution guarantees to be provided to applications without having to trust the underlying operating system [11]. XOM, however, suffers from replay attacks in which data in a compartment can be replaced by old data from that compartment. The memory integrity scheme described in this paper can complement the XOM architecture to efficiently provide tamperproof memory immunity from replays.

Terra [6] takes a different approach to trusted execution by providing each application a virtual machine to execute on, managed by a trusted virtual machine monitor. When seeking to verify some amount of data, Terra divides the data into blocks to avoid the high cost of hashing a large object, computes hashes of each block, and then stores the hash of these hashes into the VM descriptor, essentially creating a tree with two levels and a high branching factor. The scheme we propose in this paper can be used to replace the memory integrity scheme used in Terra with an efficient, optimal approach.

AEGIS [18] can run with a security kernel on top of the hardware, similar to Microsoft's Palladium [4], or without trusting the OS, similar to XOM [11]. The memory integrity scheme used by AEGIS is a Merkle hash tree, integrated

within the memory hierarchy [7]. AEGIS provides an efficient hardware implementation of Merkle trees by embedding the hash values in processor caches, but does not consider the optimal parameters for the hash tree. Our work can inform architects of secure processors on how to efficiently determine hash block sizes, which interact with the determination of cacheline sizes.

Other techniques have been introduced to provide memory integrity. A fractal-based approach [9] has been proposed to minimize the traversal of a Merkle hash tree; this work also takes the Merkle hash tree as a given and does not examine the selection of blocksize for the hash tree. Incremental multiset hash functions [3] have been proposed as a means to improve memory integrity verification performance through quick updates to logs in trusted storage, to be verified at a later time. This work focuses on sequences of reads and writes, and outperforms a hash tree only in the case of infrequent memory verification.

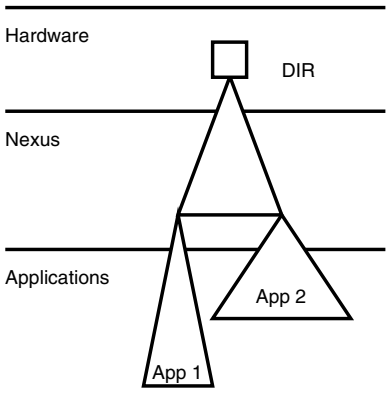

Figure 1. The Nexus provides a protected memory abstraction. Each application may have differently sized protected memory regions with different update characteristics, requiring different configurations of Merkle hash trees for efficient checking. The tree is stored in data integrity registers (DIR) in the trusted coprocessor that is part of the newly emerging TCPA standard.

3. System Model

The motivation for our work comes from the desire to develop a small trustworthy operating system that can provide applications a safe environment in which to execute. We have been building a new operating system, called Nexus, that provides attestation and secure, tamperproof execution based on the TCPA hardware (known as the TPM) [19]. While the design and implementation of this system is beyond the scope of this paper, we outline the system in order to provide a context for the use of Merkle hash trees to provide tamperproof execution.

The Nexus is a secure native operating system that provides trustworthy attestation and tamperproof execution services to its applications. It is arranged as a highly componentized system, where each component operates in a separate, isolated execution environment. The small size of the Nexus reduces the amount of code that operates with system privileges, permits the base system to be audited, and most importantly, enables the principle of least privilege to be used effectively in practice. Whereas in a monolithic operating system, all applications are dependent on, and need to trust, the implementation of all services in the kernel, Nexus applications need to trust only those components that they directly interact with. Figure 1 illustrates the structure of the Nexus.

The Nexus provides interfaces by which secure certificate chains, rooted in the *platform key* embedded in the TPM hardware, can be extended to applications. The platform key is a key embedded by the manufacturer from which other keys can be derived and using which certificate chains can be extended from the boot loader all the way to applications. This, in turn, enables the Nexus to sign certificates that say "The hardware manufacturer attests that it booted this particular version of the Nexus, which attests that it executed this version of the game client." These certificates enable remote nodes to make informed trust decisions.

As discussed before, extending trust based on a certificate requires that the system be capable of retaining predicates established at the time of certificate generation. The Nexus does this by creating a tamperproof execution environment, where the contents of memory can only be modified by the applications that have been authorized to modify them. The Nexus protects memory regions against tampering by computing a Merkle hash tree over each region and storing parts of the hash tree in the secure TPM hardware. The Nexus provides a very general interface by which applications direct the kernel to create a protected memory region of a given size, using a given blocksize. A toolkit, located in user space, is responsible for determining the optimal blocksize for the Merkle hash tree - thus, the interface is general-purpose, and the complexity of blocksize selection is left out of the kernel. The technique, shown below, is used by the toolkit to determine the optimal blocksize for the Merkle hash tree. We note that many of the other tamperproof execution schemes cited in Section 2 could use the same technique to determine the optimal blocksize in their use of Merkle hash trees.

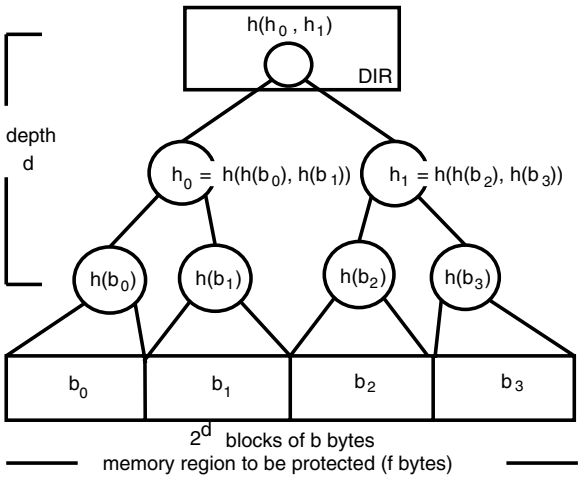

Figure 2. A Merkle tree constructed on top of f bytes of memory using a hash function h.

4. Analytical Model and Results

In this section, we derive an expression for the cost of maintaining a Merkle hash tree in the presence of uniformly distributed updates, and describe a process by which the optimal blocksize of the tree can be determined. The full derivation can be found in an accompanying technical report [21].

Without loss of generality, consider an application wishing to create and use a tamperproof memory region. We will call the size of this memory the filesize in bytes, denoted by f. We wish to divide the memory region into blocks and build a Merkle hash tree over them in order to achieve an efficient hashing based memory integrity implementation, as shown in Figure 2.

Our goal is to determine the optimal blocksize, b (in bytes), for the leaves of the tree. The Merkle tree constructed on top of the memory region is a complete binary tree, which yields the following expression, where d is the depth of the tree, that relates the blocksize to the size of the memory region and the depth of the tree:

$$f = b\,2^d$$

We assume, without loss of generality, that the memory region can be modified n times between updates to the hash tree that protects the region. n can be conservatively set to one, which will yield a data structure over which the hash tree is recomputed after every modification. In some settings, where the protected data structure in the tamperproof region is being modified as part of a transaction, there may be more than one modification between subsequent recomputations of the hash tree. The use of the n parameter captures such transactions, and $n > 1$ enables performance to be increased where timely updates to the secure summary are not necessary.

There are two components contributing to the cost of committing an update to a memory region. First, every leaf responsible for the block on which a modification is made must recompute its hash. Then, every interior node on the path from the affected leaf to the root of the tree must recompute its hash value.

4.1. Cost of Hashing: $H_l(b)$ and H_i

Each of the 2^d leaf nodes in the tree is thus responsible for computing the hash of a block of size b bytes. We model the cost of this hashing operation, referred to as the cost of a leaf hash in μseconds, by:

$$H_l(b) = \alpha b + \beta$$

Our motivation for choosing a linear model for the cost of hashing a data block of size b is based on experimental measurements of the SHA1 hash function performed on an Intel® Pentium® 4 CPU 1700MHz machine, which yielded parameters $\alpha = 0.0122348$ μsec/byte and $\beta = 1$ μsec.

In addition, each of the $2^d - 1$ interior nodes are responsible for hashing the concatenation of the values of its two child nodes. Due to the characteristics of Merkle trees, each of the child nodes contain s bytes, the size of the result of a hash operation. In the case of SHA1, $s = 20$ bytes. Thus the cost (in μsecs) of an inner hash operation is:

$$H_i = 2\alpha s + \beta$$

4.2. Leaf Hash Updates: $U_l(b)$

In order to determine the number of leaf hashes that must be recomputed after n uniformly distributed modifications to a memory region consisting of 2^d blocks, we can first consider the probability of one particular block containing a modification. The probability that the first modification is in a different block is $\frac{2^d-1}{2^d}$. We can then calculate the probability that all n modifications were in our one particular block and multiply it by the 2^d blocks each having an equal probability of being touched, leaving us with the expected number of leaf hashes that need to be recomputed after n modifications, or:

$$U_l(b) = 2^d \left(1 - \left(\frac{2^d - 1}{2^d}\right)^n\right)$$

Recall that $f = b\,2^d$, allowing each d appearing in the right side of the equation to be written in terms of b as $\log_2(\frac{f}{b})$. This substitution has been omitted for clarity.

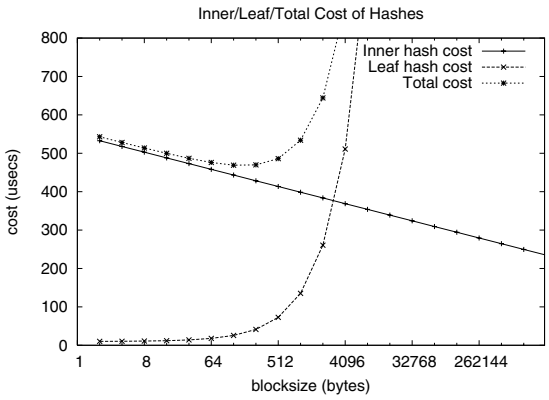

Figure 3. The relationship between the total cost of inner and leaf hashes when the number of updates is $n = 10$, and the filesize is $f = 2^{40}$ bytes. Notice the $\log$ scale on the x axis (blocksize).

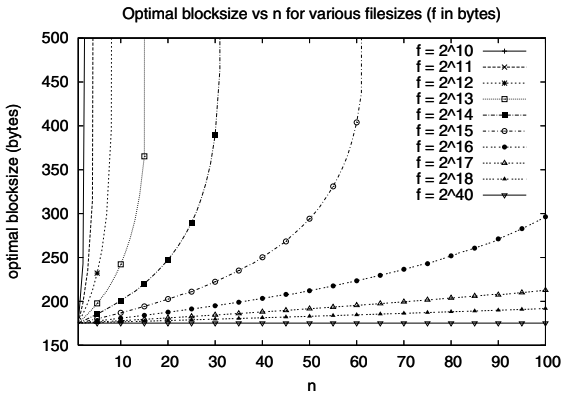

Figure 4. Optimal blocksize in bytes for a variety of file sizes f (bytes) and update intervals n using hash parameters $\alpha = 0.0122348$ μsec/byte, $\beta = 1$ μsec and $s = 20$ bytes.

4.3. Inner Hash Updates: $U_i(b)$

The expected number of inner hashes can be computed in a similar fashion to the leaf hashes. First we consider the inner nodes comprised of the immediate parents of the leaf nodes. We can think of each inner node on this level responsible for a "block" twice the size of the original blocks, one for each of the memory regions covered by the two child leaf nodes. Then the familiar leaf hash formula will apply and we can see that the number of inner hashes on the lowest level of the tree is given by substituting $d-1$ for d in that formula. A similar argument applies all the way to the root of the tree, so we can write the total number of expected inner hashes as:

$$U_i(b) = \sum_{i=0}^{d-1} 2^i \left(1 - \left(\frac{2^i - 1}{2^i}\right)^n\right)$$

The equation can be rewritten as a log term plus a constant plus some other fractions raised to the power d. In order to make our formula simple, and yet still get an accurate approximation for the number of inner hashes in the tree, we make the observation that the region we are concerned with is one in which d is relatively large. Thus these additional terms become largely unimportant, and we can write the formula for the number of inner hashes as:

$$\overline{U_i}(b) \approx nd - \sum_{i=2}^{n} (-1)^i \binom{n}{i} \left(\frac{1}{1 - \frac{1}{2^{i-1}}}\right)$$

Notice that this can be rewritten in terms of b by substituting $\log_2\left(\frac{f}{b}\right)$ for d.

4.4. Minimizing Cost

The total cost of updates to a Merkle hash tree after n modifications can be computed by combining the formulas for the expected number of leaf hashes and updates to inner nodes. This yields the following:

$$C(b) = \overline{U_i}(b)H_i + U_l(b)H_l(b)$$

Solving the derivative of this function with respect to b set equal to zero can yield the critical points, but the expression does not lend itself readily to an analytical solution. The roots can be determined, however, using a numerical method. We use a well known numerical method, Newton's method, to find the roots of this equation.

4.5. Results and Implications

The optimal blocksize is dependent on the size of the memory region to be protected (f) and the number of updates (n) to the protected region before the hash tree is updated. Figure 4 illustrates that the optimal blocksize has an asymptote at a constant fraction of the filesize f after which point the optimal blocksize is equal to f, the total filesize. Intuitively, this is the point at which enough blocks have been updated that the overhead of the inner hashes is not worthwhile, and it is faster to collapse the Merkle hash tree down to a single leaf hash over the whole region.

A natural tendency when constructing a Merkle tree is to use architectural constants, such as the native page size or the processor cache line size, for the block size. Figure 4 shows that such quantities often lead to inefficient choices. On our platform, the optimal blocksize is much less than the typical page size for large files, while it is much greater

than the typical cache line size for small files. Using the expressions derived above, it is possible to determine the optimum precisely and pick the block size that minimizes the cost of updates to the Merkle hash tree.

5. Summary

We have considered the problem of implementing a tamperproof memory region abstraction through the use of one-way hash functions. We examine the use of Merkle hash trees, whose simplicity makes them ideally suited for an efficient implementation. Yet the choice of natural parameters for hash trees, such as the native cacheline size or the native page size, often lead to inefficiencies and excessive costs when recomputing the Merkle hash tree.

This paper analytically derives an expression for the cost of updating the tree, shows that there is an optimal size given a particular combination of filesize and number of memory locations affected by each transaction, and develops a numerical technique for finding the blocksize that optimizes the cost of maintaining the tree. This work is directly applicable to the design of operating system mechanisms, as well as hardware techniques, for providing tamperproof memory. We hope that an analysis of the optimal parameter selection for increasingly ubiquitous Merkle hash trees will enable the newly available trusted hardware to be used to its full potential.

Acknowledgements

We would like to thank Fred B. Schneider for encouraging us to consider a flexible, general-purpose interface for creating protected memory regions.

References

[1] W. A. Arbaugh, D. J. Farber, and J. M. Smith. A Secure and Reliable Bootstrap Architecture. In *Proceedings of the IEEE Symposium on Security and Privacy*, pages 65–71, May 1997.

[2] M. Blum, W. Evans, P. Gemmell, S. Kannan, and M. Naor. Checking the Correctness of Memories. In *Proceedings of the 32nd Annual Symposium on Foundations of Computer Science*, pages 90–99. IEEE Computer Society Press, 1991.

[3] D. Clarke, S. Devadas, B. Gassend, M. van Dijk, and E. Suh. Incremental Multiset Hashes and their Application to Integrity Checking. In *Proceedings of the ASIACRYPT 2003 Conference*, Nov. 2003.

[4] P. England, B. Lampson, J. Manferdelli, M. Peinado, and B. Willman. A Trusted Open Platform. *Computer*, 36(7):55–62, July 2003.

[5] K. Fu, M. F. Kaashoek, and D. Mazieres. Fast and Secure Distributed Read-only File System. *ACM Transactions on Computer Systems*, 20(1):1–24, 2002.

[6] T. Garfinkel, B. Pfaff, J. Chow, M. Rosenblum, and D. Boneh. Terra: A Virtual Machine-Based Platform for Trusted Computing. In *Proceedings of the 19th Symposium on Operating System Principles*, Oct. 2003.

[7] B. Gassend, D. Clarke, G. E. Suh, M. van Dijk, and S. Devadas. Caches and Hash Trees for Efficient Memory Integrity Verification. In *Proceedings of the Ninth International Symposium on High Performance Computer Architecture (HPCA-9)*, February 2003.

[8] M. Gasser, A. Goldstein, C. Kaufman, and B. Lampson. Distributed System Security Architecture. In *Proceedings of the 12th NIST-NCSC National Computer Security Conference*, pages 305–319, 1989.

[9] M. Jakobsson, T. Leighton, S. Micali, and M. Szydlo. Fractal Merkle Tree Representation and Traversal. In *RSA-CT*, 2003.

[10] B. Lampson, M. Abadi, M. Burrows, and E. Wobber. Authentication in Distributed Systems: Theory and Practice. *ACM Transactions on Computer Systems*, 10(4):265–310, 1992.

[11] D. Lie, C. A. Thekkath, and M. Horowitz. Implementing an untrusted operating system on trusted hardware. In *Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles*, pages 178–192. ACM Press, 2003.

[12] D. Lie, C. A. Thekkath, M. Mitchell, P. Lincoln, D. Boneh, J. C. Mitchell, and M. Horowitz. Architectural Support for Copy and Tamper Resistant Software. In *Proceedings of Symposium on Architectural Support for Programming Languages and Operating Systems*, pages 168–177, 2000.

[13] U. Maheshwari, R. Vingralek, and W. Shapiro. How to Build a Trusted Database System on Untrusted Storage. In *Proceedings of the 4th Symposium on Operating Systems Design and Implementation*, pages 135–150, 2000.

[14] R. C. Merkle. Protocols for Public Key Cryptosystems. In *IEEE Symposium on Security and Privacy*, pages 122–134, 1980.

[15] R. C. Merkle. A Certified Digital Signature. In *Proceedings on Advances in cryptology*, pages 218–238. Springer-Verlag New York, Inc., 1989.

[16] W. Shapiro and R. Vingralek. How to Manage Persistent State in DRM Systems. In *Digital Rights Management Workshop*, pages 176–191, 2001.

[17] S. W. Smith and S. Weingart. Building a High-Performance, Programmable Secure Coprocessor, April 1999.

[18] G. Suh, D. Clarke, B. Gassend, M. van Dijk, and S. Devadas. Aegis: Architecture for Tamper-evident and Tamper-resistant Processing, June 2003.

[19] TCPA. Main Specification, Version 1.1a, November 2001.

[20] J. Tygar and B. Yee. Dyad: A System for Using Physically Secure Coprocessors. Technical Report CMU–CS–91–140R, Carnegie Mellon University, May 1991.

[21] D. Williams and E. G. Sirer. Optimal Parameter Selection for Efficient Memory Integrity Verification Using Merkle Hash Trees. Technical Report TR2004-1944, Cornell University Computer Science, 2004.

[22] C. K. Wong and S. S. Lam. Digital Signatures for Flows and Multicasts. *IEEE/ACM Transactions on Networking*, 7(4):502–513, 1999.

[23] B. Yee. *Using Secure Coprocessors*. PhD thesis, Carnegie Mellon University, May 1994.