IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

Distributed Bonsai Merkle Tree

Ofir Shwartz and Yitzhak Birk

CCIT Report #914
August 2017

THE ANDREW & ERNA VITERBI FACULTY OF ELECTRICAL ENGINEERING
TECHNION—ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 3200003, ISRAEL

CCIT Report #914 August 2017

Distributed Bonsai Merkle Tree

Ofir Shwartz
Electrical Engineering Dept., Technion, Israel
ofirshw@tx.technion.ac.il

Yitzhak Birk
Electrical Engineering Dept., Technion, Israel
birk@ee.technion.ac.il

Abstract

Ensuring the correct execution of a program while running on untrusted computing services is challenging. Specifically, protecting the integrity of the memory content against replay attacks while running on a platform with an untrusted memory requires dedicated tracking structures and an in-chip state. For that, the Bonsai Merkle Tree (BMT)was suggested as an efficient integrity tree. We present the Distributed Bonsai Merkle Tree, a multi-node version of the BMT suitable for parallel and distributed environments.

1. Introduction

Secure computing on untrusted environments is an emerging requirement. In these environments, it is commonly assumed that the CPU is the only trusted element, leaving everything else (including the board and the off-CPU memory) untrusted. The use of encryption enables protecting the confidentiality of the data while it resides in the untrusted memory, and the use of message authenticating code (MAC) enables protecting against forged or mis-located data; however, replay attacks, wherein old data is maliciously restored (e.g. by blocking the memory 'write' port) requires additional treatment.

The Bonsai Merkle Tree (BMT) [1] was suggested for protecting a program running on a single-node setting; however, most of the workloads that benefit from the use of untrusted environments (e.g., public clouds) are parallel and distributed in nature. Multi-node applications commonly use distributed shared memory (DSM) [2] or message passing interface (MPI) [3], where in the latter the memory space itself is not distributed, so the single-node integrity solutions are suitable. However, DSM is easier to program by simply spawning threads that access the shared address space, therefore an integrity pre-serving mechanism that supports DSM is required.

In this work we present the Distributed Bonsai Merkle Tree (DBMT). While extending the single node BMT's functionality, it does not require additional overhead on top of it.

2. Bonsai Merkle Tree

Bonsai Merkle Tree (BMT) [1] is an efficient integrity hash tree. BMT targets systems that protect their memory using counter mode encryption, wherein each memory block has its corresponding counter value. [1]'s observation is that instead of protecting the actual memory blocks using a secure hash tree, protecting the counters by a Merkle hash tree [4] (with an in-chip root hash) and using a per-block secure MAC will result in a smaller hash tree, the Bonsai Merkle Tree, which provides the same security guarantees as the original Merkle Tree. The smaller hash tree footprint results in better cache hit-rate, and therefore better performance than the Merkle Tree.

Each memory block has a small MAC alongside the data, so when it is fetched into the cache and decrypted by the counter mode technique (assuming a correct counter), forged data will result in a mismatch between the fetched MAC and the one computed over the fetched, decrypted block. Forging a counter (e.g. old counter with old data and MAC) will be detected on counter block fetch, since BMT directly protects the counter blocks (similarly to the way Merkle Tree protects the data blocks).

The BMT values are stored in the clear in the unprotected memory, and can be cached in the chip.

BMT cannot be simply used for protecting a distributed program, because its memory space spans multiple memories that are controlled by different CPUs.

3. Distributed Bonsai Merkle Tree

We assume the existence of a secure node-to-node data transfer method, such as SDSM [5].

In Distributed Bonsai Merkle Tree, we use private per-CPU encryption counters (not shared), so other CPUs cannot access them. We choose the counter corresponding to the data block to either contain the actual counter for an existing block, or NA for unmapped or a block currently resides in another CPU's local memory. Since counters are kept in blocks, counters of existing and non-existing blocks may reside in the same counter block. A per-CPU local-BMT is maintained normally, and each CPU maintains a local root hash of its local-BMT. See Fig. 1 for example.

When a memory block is needed, its counter is fetched first. If the counter value is NA, then it is considered not existing

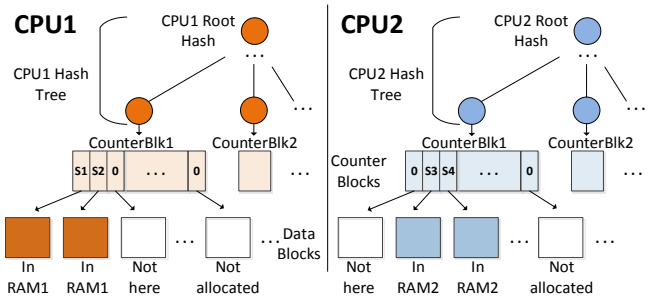

Fig. 1. DBMT of blocks of multiple states: only in CPU1, shared between CPU1 and CPU2, and not allocated.

locally, and it is then requested from the directory. When a memory block arrives (securely) in the cache from another CPU, it does not have a local counter value (NA). Only when locally evicted, a new counter value is assigned and stored locally in the private counter block, and a MAC is calculated alongside the block. When a memory block is sent and invalidated, its local counter is again reset to NA.

Any alteration of a memory block while in a CPU's local off-chip (unprotected) memory will be detected as soon as it is fetched into the cache using its MAC, and any alteration of a local encryption counter will be detected at the same time using the local-BMT.

Per-CPU encryption counters are advantageous over shared (among CPUs) counters:

1) Private counter blocks never require write permission from the directory, or value updates from remote CPUs, saving significant performance overheads for locks and communication overheads for permission re-quests.

2) Using shared counters, although a data block may only get modified by one CPU at any given time, different data blocks whose counters reside in the same counter block may get modified simultaneously. Therefore, a CPU needs exclusive write permissions for updating the counter block, so other copies at the other CPUs must be invalidated; therefore, shared counters may exhibit forced evictions although data blocks are not shared. In contrast, private counters are never shared among different CPUs, obviating the abovementioned false evictions.

Performance Analysis

Accessing locally existing blocks is similar to the case of a single-node BMT, so its performance is the same. Accessing a block that does not exist locally require bringing its counter, determining its state, and then requesting it from a remote node into the local cache. This block will not get a local counter update until it is evicted locally, and only then a DBMT update is required. Therefore, there is no performance overhead for a non-existing block, since determining the block's state is required in any DSM system, and once the block has arrived it is similar to the single-node system.

4. Conclusions

We presented the Distributed Bonsai Merkle Tree, an integrity tree suitable for multi-node and parallel environments. While extending the single-node Bonsai Merkle Tree into distributed environments, DBMT suffers from no additional overheads.

References

[1] B. Rogers, S. Chhabra, Y. Solihin, and M. Prvulovic, "Using address independent seed encryption and bonsai merkle trees to make secure processors OS-and performance-friendly", *in MICRO*'07, 2007.

[2] B. Nitzberg and V. Lo, "Distributed Shared Memory: A Survey of Issues and Algorithms," Computer (Long. Beach. Calif)., vol. 24, no. 8, pp. 52–60, 1991.D. Lie, C. Thekkath, M. Mitchell, P. Lincoln, D. Boneh, J. Mitchell and M. Horowitz, "Architectural support for copy and tamper resistant software," *ACM SIGPLAN Notices*, 2000.

[3] W. Gropp, E. Lusk, and A. Skjellum, Using MPI: Portable parallel programming with the message-passing interface, vol. 40, no. 2–3. 2000.

[4] R. C. Merkle, "A Digital Signature Based on a Conventional Encryption Function," Advances, vol. 293. pp. 369–378, 1988.

[5] O. Shwartz and Y. Birk, "SDSM: Fast and scalable security support for directory-based distributed shared memory," *Proc. IEEE Int. Symp. Hardw. Oriented Secur. Trust. HOST*, 2016.