CoSMIX: A Compiler-based System for Secure Memory Instrumentation and Execution in Enclaves

Meni Orenbach
Technion
Yan Michalevsky
Anjuna Security
Christof Fetzer
TU Dresden
Mark Silberstein
Technion

Abstract

Hardware secure enclaves are increasingly used to run complex applications. Unfortunately, existing and emerging enclave architectures do not allow secure and efficient implementation of custom page fault handlers. This limitation impedes in-enclave use of secure memory-mapped files and prevents extensions of the application memory layer commonly used in untrusted systems, such as transparent memory compression or access to remote memory.

CoSMIX is a Compiler-based system for Secure Memory Instrumentation and Execution of applications in secure enclaves. A novel memory store abstraction allows implementation of application-level secure page fault handlers that are invoked by a lightweight enclave runtime. The CoSMIX compiler instruments the application memory accesses to use one or more memory stores, guided by a global instrumentation policy or code annotations without changing application code.

The CoSMIX prototype runs on Intel SGX and is compatible with popular SGX execution environments, including SCONE and Graphene. Our evaluation of several production applications shows how CoSMIX improves their security and performance by recompiling them with appropriate memory stores. For example, unmodified Redis and Memcached key-value stores achieve about 2× speedup by using a self-paging memory store while working on datasets up to 6× larger than the enclave’s secure memory. Similarly, annotating a single line of code in a biometric verification server changes it to store its sensitive data in Oblivious RAM and makes it resilient against SGX side-channel attacks.

1 Introduction

Virtual Memory is integral to modern processor architectures. In addition to its primary role in physical memory management, it empowers developers to extend the standard memory layer with custom data storage mechanisms in software. For example, the memory-mapped file abstraction, which is broadly used, e.g., in databases [10, 5], relies on the OS’s page fault handler to map a frame and populate it with the contents of a file. Replacing accesses to physical memory with file accesses requires no application code changes. Therefore, the ability to override page fault behavior has been essential for implementing a range of system services, such as memory compression [44], disaggregation [39, 74], distributed shared memory [36, 46] and heterogeneous memory support [37].

With the emergence of Software Guard Extensions (SGX) for Trusted Execution in Intel CPUs [16, 54], applications are increasingly ported to be entirely executed in hardware-enforced enclaves [57, 45, 23, 25]. The enclave hardware protects them from attacks by a powerful privileged adversary, such as a malicious OS or a hypervisor. A number of recent systems facilitate the porting to SGX by shielding unmodified applications in an enclave [21, 80, 18]. Unfortunately, these systems do not allow secure overriding of page fault handling in enclave applications. This drawback complicates porting a large class of applications that use memory-mapped files to SGX. Further, it prevents SGX applications from using security and performance enhancements, such as efficient memory paging [60] and Oblivious RAM (ORAM) side-channel protection [66, 11, 86] without intrusive application modifications. Our goal is to eliminate these constraints.

For example, consider the task of running an SQLite database that uses memory-mapped files in the enclave. The database file must be encrypted to ensure data confidentiality. Enabling in-enclave execution of SQLite therefore requires support for encrypted memory-mapped files, which in turn implies that the page fault handler must be executed securely as well. Unfortunately, hardware enclaves available today do not support secure page faults. Instead, existing solutions use workarounds, such as eagerly reading and decrypting the whole mapped file region into trusted enclave memory [18]. This solution does not scale to large files and lacks the performance benefits of on-demand data access.

We argue that the problem is rooted in the fundamental limitation of the current SGX architecture, which does not provide the mechanism to define secure page fault handlers. Moreover, we observe that existing and emerging secure enclave architectures [28, 4, 34] suffer from similar limitations (§2).

In this work, we build CoSMIX, a compiler and a lightweight enclave runtime that overcomes the SGX architectural limitations and enables secure and efficient extensions to the memory layer of unmodified applications running in enclaves. We introduce a memory store, (mstore), a programming abstraction for implementing custom memory management extensions for enclaves. The CoSMIX compiler automatically instruments application code to allocate the selected variables and memory buffers in the mstore, replacing the ac-
cesses to enclave memory with the accesses to the mstore. The mstore logic runs in the enclave as part of the application. The CoSMIX runtime securely invokes the mstore memory management callbacks, which include custom page fault handlers. The page faults are semantically equivalent to hardware page faults yet are triggered by the CoSMIX runtime.

An mstore can implement the missing functionalities that require secure page fault handlers. For example, it may provide the secure mmap functionality by implementing the page fault handler that accesses the file and decrypts it into the application buffer. A more advanced mstore may add its own in-memory cache analogous to an OS page cache, to avoid costly accesses to the underlying storage layer. CoSMIX supports several types of mstores, adjusting the runtime to handle different mstore behaviors while optimizing the performance.

CoSMIX allows the use of multiple mstores in the same program. This can be used, for example, to leverage both secure mmap mstore and an ORAM mstore for side-channel protection. Additionally, CoSMIX supports stacking of multiple mstores to enable their efficient composition and reuse. We design and prototype three sophisticated mstores in §3.2.5, and demonstrate the benefits of stacking in §4.5.

CoSMIX’s design focuses on two primary goals: (1) minimizing the application modifications to use mstores and (2) reducing the instrumentation performance overheads. We introduce the following mechanisms to achieve them:

**Automatic inference of pointer types.** CoSMIX does not require annotating every access to a pointer. Instead, it uses inter-procedural pointer analysis [17] to determine the type of the mstore (or plain memory) to use for each pointer. When the static analysis is inconclusive, CoSMIX uses tagged pointers [47, 73, 13] with the mstore type encoded in the unused Most Significant Bits, enabling runtime detection (§3.3.1).

**Locality-optimized translation caching.** The mstore callbacks interpose on memory accesses, which are part of the performance-critical path of the application. To reduce the associated overheads, we employ static compiler optimizations to reduce the number of runtime pointer type checks and mstore accesses. These include loop transformations and a software Translation Lookaside Buffer (TLB) (§3.3.4). These mechanisms reduce the instrumentation overheads by up to two orders of magnitude (§4.2).

Our prototype targets existing SGX hardware and is compatible with several frameworks for running unmodified applications in enclaves [80, 1, 18]. However, CoSMIX makes no assumptions about enclave hardware. The CoSMIX compiler is implemented as an extension of the LLVM framework [48].

We prototype three mstores: Secure User Virtual Memory (SUVM) for efficient paging in memory-intensive applications [60], Oblivious RAM [77] for controlled side-channel protection, and a secure mmap mstore that supports access to encrypted/integrity-protected memory-mapped files. We evaluate CoSMIX on the Phoenix benchmark suite [65], as well as on unmodified production servers: memcached, Redis, SQLite, and a biometric verification server [60]. The compiler is able to correctly instrument all of these applications, some with hundreds of thousands of lines of code (LOC), without the need to manually change the application code.

Our microbenchmarks using Phoenix with SUVM and secure mmap mstores show that CoSMIX instrumentation results in a low geometric mean overhead of 20%.

For the end-to-end evaluation, we run memcached and Redis key value stores on 600 MB datasets — each about 6× the size of the secure physical memory available to SGX enclaves. In this setting, SGX hardware paging significantly affects the performance. The SUVM [60] mstore aims to optimize exactly this scenario. To use it, we only annotate the item allocator in memcached (a single line of code) and compile it with CoSMIX. Redis is compiled without adding annotations. The instrumented versions of both servers achieve about 2× speedup compared to their respective vanilla SGX baselines.

In another experiment, we evaluate a biometric verification server with a database storing 256 MB of sensitive data. We use the ORAM mstore to protect it from SGX controlled side-channel attacks [85] that may leak sensitive database access statistics. We annotate the buffers containing this database (one line of code) to use ORAM. The resulting ORAM-enhanced application provides security guarantees similar to other ORAM systems for SGX, such as ZeroTrace [66], yet without modifying the application source code. ORAM systems are known to result in dramatic performance penalties of several orders of magnitude [26]. However, our hardened application is only 5.8× slower than the vanilla SGX thanks to the benefits of selective instrumentation enabled by CoSMIX.

To summarize, our contributions are as follows:

- Design of a compiler and an mstore abstraction for transparent secure memory instrumentation (§3.2).
- Loop transformation and loop-optimized caching techniques to reduce the instrumentation overheads (§3.3.4).
- Seamless security and performance enhancement for unmodified real-world applications running in SGX, by enhancing them with the SUVM, ORAM and secure mmap mstores (§4).

## 2 Motivation

Enabling the use of custom page fault (PF) handlers in enclaves would not only facilitate porting of existing applications that rely on such functionality, but also enable a range of unique application scenarios, as we discuss next.

**SUVM.** The authors of Eleos [60] proposed Software User-space Virtual Memory (SUVM), which implements exit-less memory paging in enclaves and significantly improves the performance of memory-demanding secure applications. It keeps the page cache in the enclave’s trusted memory, while the storage layer resides in untrusted memory whose contents are encrypted and integrity-protected.
ORAM. Oblivious RAM (ORAM) obfuscates memory access patterns by shuffling the physical data locations and re-encrypting the contents upon every access. As a result, an adversary observing the accessed locations learns nothing about the actual access pattern to the data [38]. Multiple ORAM schemes have been proposed over time [38, 83, 77, 86, 66, 33], and, ORAM was recently used to manually secure applications executing in SGX enclaves against certain side-channel attacks [86, 66, 33].

Both ORAM and SUVM are generic mechanisms that could be useful in many applications. Unfortunately, integrating them with the application logic requires intrusive code modifications. With the support for efficient and secure in-enclave PF handlers, we could add these mechanisms to existing unmodified programs, as we show in the current work.

Other applications include transparent compression for in-enclave memory, mmap support for encrypted and integrity-protected files, and inter-enclave shared memory, as well as various memory-management mechanisms [39, 74, 37].

Unfortunately, existing enclave hardware provides no adequate mechanisms to implement efficient and secure user-defined PF handlers, as we describe next.

2.1 Background: page-faults in enclaves

There are several leading enclave architectures: Intel SGX [16, 43], Komodo for ARM Trust Zone [34, 15], Sanctum [28], and Keystone [4]. Among these, only Intel SGX and Sanctum published support for paging. We briefly describe them below.

**Intel SGX [16, 43, 54]** supports on-demand paging between secure and untrusted memory. SGX relies on Virtual Memory hardware in X86 CPUs. When a PF occurs, the enclave exits to an untrusted privileged OS which invokes the SGX PF handler. The enclave execution resumes (via ERESUME) after the swapping is complete.

Since the PF handler is untrusted, the SGX paging is secured via SGX paging instructions. Specifically, EWB encrypts and signs the page when swapping the page out, whereas ELDU validates the integrity and decrypts when swapping it in. These instructions cannot be modified to perform other operations. They cannot change the internal SGX encryption key or modify the swapped page. In other words, they cannot act as a general-purpose secure PF handler.

**Sanctum** [28] supports per-enclave page tables and secure PF handlers. It uses a security monitor that runs at a higher privilege level than the OS and the hypervisor. Upon a PF, the enclave exits to the security monitor, which triggers the in-enclave secure PF handler.

2.2 Limitations of existing enclaves

**Signal handling in SGX.** Page fault handlers can be customized in userspace by registering a signal handler. SGX supports signal handlers in enclaves and works according to the following scheme (Figure 1): when an interrupt occurs, the enclave exits to the OS. The OS takes control and performs an up-call to an untrusted user-space trampoline in the enclave’s hosting process. The trampoline re-enters the enclave by invoking the in-enclave signal handler. After the signal handler terminates, the enclave exits and resumes execution of the last instruction via ERESUME.

**SGX: No secure page fault handlers.** The SGX signal handling mechanism cannot guarantee secure execution of the handler itself. When the enclave is resumed after the PF, ERESUME replays the faulting memory access. Therefore, the enclave cannot validate that the signal handler was indeed executed correctly, or was executed at all. It remains to be seen whether the next version of SGX will solve this problem.

**SGX: Performance overheads.** Even with hardware support for the secure signal handler, SGX has inherent architectural properties that will lead to significant performance penalties, rendering this mechanism unsuitable for customized application memory management. The architecture relies on the OS to manage the enclave’s virtual memory. Furthermore, SGX may only run userspace code. Since the OS is untrusted, any secure page fault handling would inevitably follow the signal handling scheme depicted in Figure 1, namely, double enclave transition between the trusted and untrusted contexts.

We measure the latency of an empty SGX PF handler (access to a protected page) to be 11μsec, which is more than 6× the latency of a signal handler outside SGX. For comparison, CoSMIX’s software page fault handler is only 0.01 μsecond, the cost of a single function call, which is three orders of magnitude faster than in SGX.

Further analysis shows that the signal latency is dominated by the latency of enclave transitions, which we measure to be 5.3μsec each 1 and stems from costly validation, register scrubbing, TLB and L1 cache flushes [81, 43].

**Other enclave architectures.** Enclave transition overheads are pertinent to other enclave architectures. Komodo reports exit latency of 0.96μsec [34]. Sanctum and Keystone do not disclose their enclave transition penalties, yet they describe similar operations performed when such transitions occur.

We conclude that secure page fault handlers are not supported in SGX, and are likely to incur high-performance costs in other enclave architectures due to transition overheads.

---

1This value is almost double the one reported in prior works [60, 84] because of the firmware update to mitigate the Foreshadow [81] bug.
2.3 Code instrumentation for enclaves

Instrumenting application memory accesses with the desired software PF handling logic is a viable option to achieve the functionality equivalent to the hardware-triggered PF handlers. Unfortunately, existing instrumentation techniques are not sufficient to achieve our goals, as we discuss below.

**Binary instrumentation.** Dynamic binary instrumentation tools [51, 56, 24, 52, 70], such as Intel PIN [51], enable instrumentation of memory accesses. Unfortunately, these tools have significant drawbacks in the context of in-enclave execution. For example, for PIN to work, all its libraries should be included in the enclave’s Trusted Computing Base (TCB). Moreover, PIN requires JIT-execution mode to instrument memory accesses. Therefore, the enclave code pages should be configured as writable, which might lead to security vulnerabilities. Removing the write access permission from enclave pages will be supported in SGX V2, but doing so will require costly enclave transitions [43].

Static binary instrumentation tools do not suffer from these shortcomings. However, compared to the compiler-based techniques we propose in this work, they do not allow using comprehensive code analysis necessary for performance optimizations. Therefore, we decided against the binary instrumentation design.

**Compiler-based instrumentation.** The main advantage of this method is the ability to aggressively reduce the instrumentation overheads by using advanced source code analysis. On the other hand, the source code access requirement limits the applicability of this method. However, this drawback is less critical in the case of SGX enclaves because many SGX execution frameworks, such as Panoply [76] and SCON [18], require code recompilation anyway. Therefore, we opt for compiler-based instrumentation in CoSMIX.

3 CoSMIX Design

CoSMIX aims to facilitate the integration of different mstores efficiently into SGX applications. Our design goals are:

- **Performance.** Low overhead memory-access and software address translation.
- **Ease-of-use.** Annotation-based or automatic instrumentation without manual application code modification.
- **General memory extension interface.** Easy and modular development of application-specific memory instrumentation libraries.
- **Security.** Keep SGX security guarantees and small TCB.

**Threat Model.** CoSMIX is designed with the SGX threat model, where the TCB includes the processor package and the enclave’s code. Additionally, we assume that the code running in an enclave does not contain memory vulnerabilities.

![CoSMIX compilation overview. The compiler is guided by code annotations and global instrumentation policy.](image)

3.1 Design overview

**Compiler-based instrumentation.** CoSMIX enables SGX developers to build custom memory stores, mstores, that redefine the memory layer functionality of an instrumented program. To integrate one or more mstores into an application, the CoSMIX compiler automatically instruments the program (Figure 2). The developer may selectively annotate static variables or memory allocations to use different mstores, or define a global instrumentation policy. The compiler automatically instruments the necessary subset of memory accesses with the accesses to the corresponding mstores, and statically links mstore libraries with the code.

The CoSMIX configuration file defines the instrumentation behavior. It specifies annotation symbols per mstore, mstore type (§3.2) and the instrumentation policy(§3.3).

3.2 Mstore abstraction

At a high level, an mstore implements another layer of virtual memory on top of an abstract storage layer. An mstore operates on pages, mpages, and keeps track of the mpage-to-data mappings in its internal mpage table. When an application accesses memory, the runtime invokes the mstore’s software page fault handler, retrieves the contents (e.g., for the secure mmap mstore, it would read data from a file and decrypt), and makes it accessible to the application.

We distinguish between cached and direct-access mstores. A cached mstore maintains its own mpage cache to reduce accesses to the storage layer, whereas a direct-access mstore does not cache the contents.

Figure 3 shows the execution of an access to a cached mstore. The pointer access triggers the runtime call, which chooses the appropriate mstore and checks the translation cache. If the translation is not in the cache, the runtime invokes the mpage fault handler. The mstore translates the pointer, and either fetches the referenced mpage from the page cache, or retrieves it from the storage layer and updates the mpage and translation caches.

3.2.1 Mstore callbacks

Table 1 lists the callback functions mstores must implement.

**Initialization/teardown.** The mstore is initialized at the beginning of the program execution and torn down when the pro-
3.2.2 Pointer access and mpage faults

When the instrumented code accesses the mstore, the runtime incurs an equivalent of a page fault, and invokes the respective callback in the mstore, as discussed below.

Cached mstores. A cached mstore translates the pointer to the mpage inside its cache. mpf_handler_c returns the pointer into the mpage that holds the requested data, allowing direct access from the code. For cross-page misaligned accesses, the runtime creates a temporary buffer and populates it by calling mpf_handler_c for every page separately. For store access, the updates are written to both pages.

When the runtime determines that the code is accessing the same mpage multiple times, it may cache the pointer to that mpage in its private TLB. This avoids the address translation overheads for all but the first access to the page. To ensure that the mpage is not swapped out by the mstore logic, the runtime notifies the mstore to pin the mpage in the mpage cache via notify_tlb_cached. The page is unpinned by notify_tlb_dropped when its translation is evicted from the TLB (§3.3.4 for more details).

There can be multiple cached mstores in the same program, each with its own mpage size. The runtime can query an mstore page size using the get_mpage_size call, for example, to determine accesses to the same mpage in the TLB.

A cached mstore must implement the flush() callback to synchronize its mpage cache with the storage layer. This callback is used, for example, to implement the msync call.

Direct-access mstores. Direct-access mstores have no cache and thus are easier to implement. The input pointer provided to the mpf_handler_d callback may be used without address translation, and at finer granularity not bound to the mpage size. The runtime provides a thread-local intermediate buffer to store the accessed data. For loads, the program uses this buffer. For stores, the runtime writes the updated contents back to the mstore using the write_back callback.

3.2.3 Thread safety and memory consistency

CoSMIX allows multiple threads to access the same mstore, as long as the mstore implementation is thread-safe.

For cached mstores, CoSMIX does not change the mstore memory consistency model as long as the accesses are inside the same mpage, and the mpage size is 4KB or larger. The CoSMIX runtime does not introduce extra memory copies for such accesses and effectively inherits the mstore memory consistency. In addition, the mstore itself must ensure that the storage layer is accessed only via its mpage cache, thereby preventing different threads from seeing inconsistent data.

This guarantee does not hold for direct-access mstores and misaligned cross-mpages. The primary implication is that hardware atomics will not work for such accesses.

We believe that this limitation does not affect most practical cases. The lack of cross-mpage atomics support does not affect race-free programs synchronized via locks. This is because the intermediate buffer is written back to the mstore immediately after the original store operation and thus will be protected by the original lock. We observed no cases of cross-mpage atomics in the evaluated applications.
We support private file mapping only and leave cross-enclave
(§3.3). Direct-access

3.2.4 Memory vs. file-backed mstores

The mstore abstraction described so far instruments memory access
directly, but it is insufficient to enable implement-
tion of memory-mapped files. For example, consider a
program that uses mmap to access a file, and then calls fsync. It
will not work correctly because fsync is not aware of the
mstore’s internal cache. Specifically, all the I/O operations on
files managed by a file-backed mstore must interact with its
mpage cache to avoid data inconsistency.

We define a file-backed cached mstore type, which imple-
ments all the callbacks of memory-backed mstores, but
additionally overrides the POSIX file API that interacts with
the page cache, e.g., open, close, read, write, msync (§3.3). Direct-access mstores do not have internal caches;
they can be used with files without overriding file I/O
operations other than mmap itself.

3.2.5 Mstore examples

CoSMIX provides a set of reusable building blocks for
mstores, such as a slab memory allocator, spinlocks, a generic
mpage table, and an npage cache, all with multi-threading
support, which we use to implement the mstores below.

SUVM mstore. SUVM allows exit-less memory paging for
enclaves by keeping a page table, page cache and a fault
handler as part of the enclave. We implement SUVM from
scratch as a cached memory-backed mstore, using CoSMIX’s
generic mpage table and mpage cache. The alloc function
returns a pointer to the storage layer in untrusted memory.
Upon mpf_handler_c, the mstore checks whether the needed
mpage is already cached in the mpage table. If not, it reads the
mpage’s contents from the storage layer, decrypts and verifies
its integrity using a signature maintained for every mpage,
and finally copies it to the mpage in the page cache. When the
mpage cache is full, the mstore evicts pages back into the
storage layer.

Secure mmap mstore. This mstore enables the use of memory-
mapped encrypted and integrity-protected files in enclaves.
We support private file mapping only and leave cross-enclave
sharing support for future work. This is a cached file-backed
mstore that maintains its own mpage table and mpage cache.

The alloc callback is invoked by the runtime upon the
mmap call in original code. alloc records the mapping start
offset in the file and the access flags in an internal table. It
then returns a pointer with the requested offset from a unique
file base address. This address is internally allocated by the
mstore and used as isolated address space for each file.

The mpf_handler_c callback translates the given pointer
to the respective file. If the contents are not available in the
mpage cache, the data is fetched from the file using a read
system call, followed by decryption and integrity check.

Oblivious RAM mstore. ORAM obfuscates memory access
patterns by shuffling and re-encrypting the data upon every
data access. The ORAM mstore streamlines the use of ORAM
in enclaves. It allows the developer to allocate and access the
buffers that store sensitive data in an ORAM, thereby protect-
ing the program against controlled side-channel attacks [85].

We implement a direct-access memory-backed ORAM
mstore. This is because if it were cached, the accesses to
the cache would be visible to the attacker, compromising the
ORAM security guarantees. Our ORAM mstore addresses a
threat-model similar to ZeroTrace [66], yet without leakage
protection in the case of enclave shutdowns. Specifically, all
the instrumented memory accesses destined to the ORAM
mstore become oblivious, such that an adversary cannot learn
the actual memory access pattern. We implement the Path
ORAM algorithm [77] and ensure oblivious accesses to its
position map and stash using the cmovz instruction.

We store the Path-ORAM tree in a contiguous buffer within
the enclave trusted memory. This eliminates the need to im-
plement block encryption and integrity checks as part of the
ORAM mstore since SGX hardware does exactly that.

The alloc function allocates the requested number of
blocks in ORAM and registers them with the ORAM module.
It returns an address from a linear range with a hard-coded
base address, which is used only to compute the block index.

The mpf_handler_d callback translates the address to the
requested block index and invokes the ORAM algorithm to
obliviously fetch the requested memory block to a tempo-
rary buffer. Loads are issued from this buffer and stores are
appended with the write back callback.

3.2.6 Stacking mstores

The mstore abstraction makes it possible to stack different
mstores. Stacking allows only one mstore to invoke another mstore
recursively. We denote by A → B an mstore A that internally
accesses its memory via mstore B.

Consider, for example, ORAM → SUVM. The motivation
to stack is when the ORAM mstore covers a region that is
larger than the enclave’s physical memory. Since SUVM opti-
mizes the SGX paging mechanism, stacking ORAM on top
of SUVM improves ORAM’s performance (§4.5).

To create a new A → B mstore, the developer simply an-
notates A’s storage layer allocations with B’s annotations.
CoSMIX instruments all these accesses appropriately.

Stacking the ORAM mstore on top of any mstore that main-
tains data confidentiality does not compromise the ORAM
access pattern obfuscation guarantees, as ORAM protocols
consider the backing store to be untrusted [77]. Therefore
ORAM → SUVM would maintain data-oblivious access.
However, the stacking → operator is not commutative from the security perspective: SUVM→ORAM would result in the SUVM mstore caching mpages fetched obliviously by the ORAM mstore, thereby leaking the access patterns to these mpages and compromising ORAM’s security guarantees.

### 3.3 CoSMIX compiler and runtime

The instrumentation compiler modifies the application to use mstores and is guided by code annotations or/and a global instrumentation policy. The compiler needs to instrument four different types of code: (1) memory accesses; (2) memory management functions; (3) file I/O operations for file-backed mstores; (4) libc library calls.

**Instrumentation policy.** A developer may annotate any static variable declaration or memory allocation function call. Annotations allow instrumentation of a subset of the used buffers to reduce instrumentation overheads. Alternatively, a global instrumentation policy specifies compile-time rules applied to the whole code base (e.g., instrument all calls to malloc), or run-time checks injected by the compiler (e.g., using mstore for large buffers above a certain threshold). A global policy serves for bulk operations on large code bases, such as adding SUVM mstore to Redis sources with over 130K LOC (§4.4).

Similarly, for file-backed mstores, a global policy may limit the use of the mstore to specific files or directories.

#### 3.3.1 Pointer access instrumentation

**Static analysis.** The compiler uses static build. Therefore, it can conservatively determine the subset of operations that must be replaced with mstore-related functions at compile time and eliminate the instrumentation overhead for such cases. Trivially, the compiler may replace an annotated call to malloc with the alloc callback of the requested mstore. A much more challenging task, however, is to determine the type of the mstore (if any) to use for every pointer in the program.

For this purpose, we use Andersen’s analysis [17] to generate inter-procedural point-to information. In a nutshell, CoSMIX first parses all instructions and generates a graph with pointers as nodes and their constraints (e.g., assignment or copy) as edges. The graph is then processed by a constraint solver which outputs the set of points-to-information.

When instrumenting memory accesses, CoSMIX can use this information to determine whether the pointer may alias to a specific mstore pointer.

**Runtime checks and tagged pointers.** CoSMIX’s pointer analysis is sound but incomplete; therefore it requires runtime decisions for ambivalent cases. We use tagged pointers [47, 13, 73, 31] to determine pointer type at runtime. Each mstore is assigned a unique identifier, stored in unused most significant bits of the pointer virtual address. For instrumented allocations, the runtime adds this identifier to the returned address from the mstore allocation. For external function calls and memory accesses, the runtime checks the tag, strips it from the pointer, and invokes the callback of the respective mstore if necessary.

**Tagged pointers vs. range checks.** One known limitation of tagged pointers is that the application code might reset the higher bits of a pointer. Prior work [47] and our own experience suggest that this is rarely the case in practice. An alternative approach is to differentiate between mstores by assigning a unique memory range to each. Using tagged pointers turned out to be faster in our experience because the range check requires additional memory accesses.

#### 3.3.2 Memory management and file I/O calls

The compiler replaces all the memory management operations selected by the instrumentation policy with the calls to the runtime that invokes the appropriate mstore callbacks.

Similarly, file I/O operations are replaced with runtime wrappers. On open, the runtime determines whether to use an mstore with the current file and registers its file descriptor. An I/O call using this file descriptor will be redirected to the respective mstore.

#### 3.3.3 libc support

Invoking an uninstrumented function on an mstore pointer would result in undefined behavior. We assume that most application libraries are available at compile time. However, libc is not instrumented and we provide wrappers instead, similarly to other works [47, 58].

There are two main reasons to not instrument libc. First, doing so would create a bootstrapping problem since mstores might use libc calls. Second, SGX runtimes such as SCONE [18] use proprietary, closed-source versions of libc. Wrapping libc functions allows CoSMIX to be portable across multiple enclave execution frameworks.

CoSMIX provides wrappers for most popular libc functions (about 80), which suffices to run the large production applications we use in our evaluation. Adding wrappers for more functions is an ongoing effort.

In addition to stripping the pointer tag, these wrappers must guarantee access to virtually contiguous input/output buffers from uninstrumented code, instead of using mstore mpages. Thus, where necessary, the wrappers use a temporary buffer in regular memory to stage the mstore buffer before the library call and write it back to the mstore after the call.

#### 3.3.4 Translation caching

Minimizing the instrumentation overhead is a fundamental requirement for CoSMIX. The overheads are caused mainly by runtime checks of the pointer type and invocation of the mstore logic on memory accesses.

To reduce these overheads, CoSMIX first runs aggressive generic code optimizations, reducing memory accesses. It
also avoids invoking mstore page fault handlers for recurrent accesses to the same mpage.

**Opportunistic caching.** We introduce a small (one cache line) TLB stored in the thread-local memory. This TLB is checked upon each access to the mstore. The runtime pins the page in the mstore while the mpage translation is cached, and unpins when it is flushed. To support multiple threads, the TLB notification callbacks use a reference counter for each mpage. The mpage can be evicted if the counter drops to zero, eliminating the need for explicit TLB invalidation. We choose small TLB size (5) for its low lookup times. Increasing the size did not improve performance in our workloads.

**Translation caching in loops.** The TLB captures the locality of accesses quite effectively, but in loops, the performance can be further improved by transforming the code itself to use the mpage base address without checking the TLB.

For example, in the case of an array allocated in an mstore and sequentially accessed in a tight loop, most accesses to the mstore are performed within the same mpage. Therefore, replacing the code in the loop to check the TLB only at mpage boundaries would result in near-native access latency.

To perform this optimization, CoSMIX has to (a) determine the iterations in which the same mpage gets accessed, and (b) determine the pointer transformation across the iterations. For (b), CoSMIX uses the scalar evolution analysis [82] in the compiler to find predictable affine transformations for the pointers used in the loop. For (a) it injects a code that determines the number of iterations where the cached translation hits the same mpage, recomputing the new base pointer and dropping the translation from the TLB when crossing into a new mpage. Finally, it replaces the original accesses to the mstore with the accesses to the mpage’s base pointer with the offset, which is updated across the loop iterations according to the determined transformation.

### 3.4 Discussion

**Security guarantees.** CoSMIX itself does not change the security of SGX enclaves. Its runtime neither accesses untrusted memory nor leaks secret information from the enclave. However, the security of mstores depends on their implementation: SUVM has the same security as SGX paging [60] and the ORAM mstore introduces a controlled side-channel [85] protection mechanism not available in the bare-metal SGX [66]. We note, however, that CoSMIX does not guarantee that the code using an mstore will indeed maintain that mstore’s security properties. For example, in case of the ORAM mstore, user code must not use sensitive values read from ORAM in a data/control dependent manner because doing so might break the data pattern obfuscation [64].

**Other mstore applications.** Mstores are general and can be used to implement many other useful extensions. For example, implementing bounds checking for 32-bit address space enclaves as in SGXBounds [47] becomes easy. All it takes is adding an extra 4 bytes for each buffer allocation to store the lower bounds of the object and tag the pointer’s highest 32 bits with its lower bounds’ address. Then, every memory access is instrumented to check these bounds. The SGXBounds mstore can implement this logic in its callback functions.

Another useful application is transparent inter-enclave shared memory, which may enable execution of multi-socket enclaves and support for secure file sharing.

**Eleos vs. CoSMIX.** The starting point of our design was Eleos [60]. There, the authors introduce spointers, which are similar to C++ smart pointers. In Eleos all necessary memory accesses are replaced with spointers and translations are cached in the spointers themselves. On every access, spointers perform bound checking, to make sure that pointer arithmetic on the spointer did not cross to a new page. However, the mpage bound check impacts performance greatly, even when the caching is limited to the scope of a function. Second, maintaining static translation for mstore pointers complicated the design. For example, pointer-to-integer casts had to be invalidated, forcing a reverse mapping in mstores.

A key lesson from CoSMIX is that caching the translations only in cases of high access locality is enough to leverage the performance benefits and simplify the design.

**Limitations.** Inline assembly snippets, while quite rare, cannot be easily supported. CoSMIX considers them as an opaque function call. It injects code to check whether passed arguments are mstore pointers. If so, CoSMIX aborts the program and notifies that manual instrumentation is necessary.

**Hardware extensions.** We hope that CoSMIX will motivate hardware developers to support secure in-enclave fault handling. This functionality would allow enclaves to control the execution flow of the page faults. For example, an enclave might refuse to resume execution after a fault unless a correct secure fault handler has been invoked. As a result, secure page faults would allow extending enclaves with cached mstore functionality, such as the secured mmap provided by CoSMIX, albeit at a significantly higher performance costs due to transitions to/from untrusted mode. Moreover, hardware support for secure fault handlers would enable paging of code pages not supported by CoSMIX.

However, direct mstores such as ORAM cannot be supported in the same way, since they invoke the fault handler for every memory access. Therefore, good performance could be achieved only by much more intrusive modifications that would avoid enclave mode transitions. Additionally, enclave hardware evolves slowly. For example, the SGX2 specification was published in 2014, yet is still not publicly available in mainstream processors [20]. CoSMIX on the other hand, can be used to enhance enclaves’ functionality today.
4 Evaluation

Implementation. CoSMIX implementation closely follows its design. The compiler prototype is based on the LLVM 6.0 framework and is implemented as a compile-time module pass, consisting of 1,080 LOC. We compile applications using the Link-Time Optimization (LTO) feature of LLVM and pass them as inputs to the compiler pass.

CoSMIX uses the SVF framework to perform Andersen’s pointer analysis with type differentiation. CoSMIX runtime is written in C++ and consists of 1,600 LOC. CoSMIX’s configuration file is JSON formatted and CoSMIX uses Jansson [3] to parse it. All mstores are written in C++. Their implementation follows the design described in Section 3. The implementations of SUVM, ORAM and mmap mstores are 935 LOC, 551 LOC, and 1,108 LOC respectively. mmap mstore leverages the SCONE file shields, which override the read/write calls in libc to implement integrity checks and encryption for file I/O operations. This is one of the examples when our design choice to use libc wrappers rather than libc instrumentation pays off.

Setup. Our setup comprises two machines: server and client. The client generates the load for the server under evaluation. The client and the server are connected back-to-back via a 56Gb Connect-IB operating in IP over Infiniband (27) (32Gbps effective throughput) to stress the application logic and avoid network bottlenecks.

For the server, we use Dell OptiPlex 7040, with Intel Skylake i7-6700 4-core CPU with 8 MB LLC, 16 GB RAM, and 256 GB SSD drive, Ubuntu Linux 16.04 64-bit, Linux 4.10.0-42, and Intel SGX driver 2.0 [2]. We use LLVM 6.0 to compile the source code. As recommended by Intel, we apply LIITF microcode patches [9]. The client runs on a 6-core Intel Xeon CPU E5-2620 v3 at 2.40GHz with 32GB of RAM, Ubuntu 16.04 64-bit, Linux 4.4.0-139.

Methodology. Unless otherwise specified, we run all the workloads in SGX using SCONE [18]. We run each test 10 times and report the mean value, with the standard deviation below 5%. We compile all workloads as follows: (1) compile to LLVM IR with full optimizations (-O3) and invoke LLVM’s IR linker to link all IR files; (2) invoke CoSMIX LLVM pass for code instrumentation; (3) use the SCONE compiler to generate an executable binary linked with its custom libc. We skip step (2) when compiling the baseline.

Summary of workloads. We evaluated several production applications and benchmarks, detailed in Table 2. CoSMIX successfully instruments large code bases using only a few or no code annotations, and no source code changes.

4.1 Mstore performance

First, we measure the latency of random accesses to mstores and compare them to native memory accesses in the enclave. We evaluate scenarios with small and large memory usage where the latter causes SGX page faults. We report the results for SUVM and ORAM mstores and exclude the mmap mstore because it is similar to SUVM. We measure the average latency to access random 4 KB pages over 100k requests.

Table 3 shows the results. For small datasets, SUVM incurs low overhead compared to regular memory accesses. However, for large data sets which involve SGX paging, it is about 10× faster. This is because SUVM optimizes the enclave paging performance by eliminating enclave transitions [60]. As expected, ORAM mstore access is between 30× to 60× slower than the baseline, even when covering a small range of 16 MB. This result indicates that selective instrumentation for ORAM is essential to achieve practically viable systems.

4.2 Instrumentation and mstore overheads

We instrument all seven benchmarks in the Phoenix suite [65] to measure CoSMIX’s instrumentation overheads. Each benchmark is small, making the results easier to analyze.

We evaluate 4 configurations: (1) Full automatic instrumentation using SUVM and mmap mstore. Both mstores are run side-by-side because Phoenix uses both dynamic allocations and memory-mapped files. (2) Same but with an empty mstore. All the pointers are instrumented, but the mstore logic is not invoked. (3) Selective instrumentation where we manually annotate only the inputs. (4) Same but with an empty mstore. For benchmarks that use mmap, the baseline reads the entire input file to memory.

Measuring an empty mstore allows us to distinguish between overheads of pointer instrumentation and mstores.

To focus solely on the CoSMIX overheads, we use the small dataset shipped with Phoenix so no paging operation will occur. The only exception is the histogram benchmark, for which we synthetically resize the dataset to 25 MB. We exclude mstore initialization and preload all the datasets into

---

2CoSMIX source code is publicly available at https://github.com/acsl-technion/cosmix.
SGX memory, both in the baseline and in CoSMIX measurements, to stress the runtime components of the system.

Often mstores can be tuned to reduce the translation overhead, by increasing the mstore page size. As a result, the accesses in a loop might touch fewer pages, enabling more efficient use of CoSMIX’s TLB. Therefore, we manually tune the page size, setting it to 256 KB for kmeans, 16 MB for word count, 64 MB for lreg, and 4 KB for the rest of the benchmarks, as in all the other experiments.

Figure 4 shows the results. Figure 4b shows the overhead for both CoSMIX’s instrumentation and mstore logic. Figure 4a excludes the mstore logic. In each figure, the rightmost bar refers to the selective instrumentation of accesses to the input data alone, and the other bar refers to full instrumentation of all dynamic allocations and mmap calls.

Instrumentation overheads. The runtime overheads excluding mstores are relatively small, with an average (geomean) of 17% for full instrumentation and 10% for the input instrumentation alone, with the worst case of 50% in lreg.

Full instrumentation. With the full instrumentation, mstore logic dominates the runtime overheads, ranging from almost none for histogram to 26× for matrix multiplication. Such variability stems from the different ways memory is accessed. Specifically, if the program exhibits poor access locality, or the CoSMIX compiler fails to capture the existing locality in a loop, the runtime will not be able to optimize out the calls to the mstore inside the loop, resulting in high overheads.

Selective instrumentation. Instrumenting only the input buffers results in dramatically lower overheads, ranging from 5% to 15%. The only pathological case is kmeans, where the CoSMIX compiler fails to optimize accesses to the multi-dimensional input array because the inner array is too short. Unrolling this loop reduces the overhead to about 5%. We plan to add automatic optimizations for this case in the future.

CoSMIX overheads for non-enclave execution. For completeness we perform the same experiment with Phoenix but now outside the enclave. As expected, the relative overheads increase as compared to the in-enclave execution, with up to 50% geometric mean slowdown when using selective instrumentation with SUVM and mmap mstores and up to 25% when using an empty mstore. We attribute this discrepancy to SGX’s memory encryption engine [40], which provides confidentiality and integrity to memory accesses and therefore offsets the CoSMIX instrumentation overheads.

### 4.3 Secure mmap with SQLite

SQLite is a popular database, but running it in SGX with mmap while providing encryption and integrity guarantees for the accessed files is not possible today. To run SQLite with mmap support, we use the secure mmap mstore. We configure the mpage size to be 4 MB. We use SQLite v3.25.3, and evaluate it with kvtest [10], shipped with SQLite to test read access latency to DB BLOBs. We use a database stored in a 60 MB file holding 1 KB BLOBs. The database is sized to fit in SGX physical memory. This allows us to focus on the evaluation of the file access logic rather than SGX paging (refer to §4.4 for paging evaluation).

As a baseline, we evaluate SQLite with its internal backend that uses read/write calls instead of memory-mapped files. In this configuration, SQLite implements its own optimized page cache for data it reads from files. In the evaluation, we vary the SQLite page cache size from disabled (1 KB) to 60 MB (no misses). We measure the average latency of 1 KB random read requests over 1 million requests.

<table>
<thead>
<tr>
<th></th>
<th>CoSMIX</th>
<th>read</th>
<th>read</th>
<th>read</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mmap</td>
<td>cache</td>
<td>cache</td>
<td>cache</td>
</tr>
<tr>
<td>Query latency</td>
<td>2.4 µsec</td>
<td>10.7 sec</td>
<td>4.5 sec</td>
<td>1.8 sec</td>
</tr>
<tr>
<td>Speedup</td>
<td>4.4 ×</td>
<td>8 ×</td>
<td>4 − 16 ×</td>
<td>1.8 ×</td>
</tr>
</tbody>
</table>

Table 4: SQLite performance with secure mmap mstore.
4.4 Optimizing memory-intensive workloads with the SUVM mstore

To demonstrate CoSMIX’s support for different SGX execution frameworks, we run the following experiments both in SCON [18] and Anjuna [1].

We use CoSMIX to accelerate SGX execution of applications with a large memory footprint. We choose Redis and memcached key-value stores as representatives. Both run with data sets of 600 MB – about 6× larger than the SGX enclave-accessible physical memory. These applications experience a significant slowdown due to SGX paging overheads. The goal is to reduce these overheads using the SUVM mstore.

Memcached uses a slab allocator to manage its memory. We annotate a single line of code where the memory is allocated, making SUVM mstore manage all items.

We evaluate memcached v1.4.25 [35] using the memaslap load generator shipped with libmemcached v1.0.18 [6]. Our workload consists of 10% SET and 90% GET requests for 1 KB items (key+value) as used in prior works [60], with requests uniformly distributed to stress the memory subsystem.

We compare the instrumented memcached with SUVM to native SGX execution. In addition, we run a manually optimized version of memcached with the SUVM used in Eleos [60]. Notably, in Eleos, the authors changed memcached internals to create a shadow memory buffer for the slab allocator. This is an intrusive change that CoSMIX eliminates completely. All runs are performed on 4 cores. Figure 5a shows that SUVM mstore boosts the throughput by 1.9× and 2.2× compared to native SGX execution in Anjuna and SCON respectively. The difference between the frameworks correlates with the relative time each of them spends resolving page faults. Interestingly, the CoSMIX version is about 7% faster than Eleos thanks to its compiler optimizations.

Redis. Manually annotating Redis with its 130 KLOC would be too tedious. Further, its memory management involves too many allocations, making annotation challenging. Therefore, we use automatic instrumentation without code changes.

To achieve high performance, we leverage CoSMIX’s ability to perform conditional instrumentation. Specifically, CoSMIX introduces runtime checks that determine whether to allocate a buffer in an mstore or in regular memory based on the requested allocation size. In Redis, we configure the policy to redirect all allocations in the range of 1 KB-10 KB to the SUVM mstore. The intuition is to use SUVM only for keys and values and keep native memory accesses for the rest.

We use Redis v5.0.2 [8], and evaluate it using the memtier v1.2.15 official RedisLab load generator [7]. We configure memtier to generate uniformly distributed GET requests for 1 KB items as used in prior works [18].

Figure 5b shows that Redis with CoSMIX achieves about 1.6× and 2× higher throughput compared to native SGX execution in Anjuna and SCON respectively. These results demonstrate the power of CoSMIX to improve the performance of an unmodified production system.

4.5 Protecting data with the ORAM mstore

Selective instrumentation capabilities in CoSMIX are particularly useful when using heavyweight mstores such as ORAM. ORAM is known to dramatically affect performance (Table 3).

We use ORAM to protect a face verification server [60] against controlled side-channel attacks [85] on its data store.

The server mimics the behavior of a biometric border control server. It stores a database with sensitive face images. When a person passes through border control, the client at the border kiosk queries the server whether the image of the person in the database matches the one taken at the kiosk.

The implementation stores the images in an array. The server fetches the image from the array and compares it with the input image, using the LBP algorithm [12]. This implementation is vulnerable to controlled channel attacks which leak the access pattern to SGX memory pages. Thus, an attacker may observe page access frequency and learn when a person passes through border control. Note that existing defenses against controlled channel attacks would be ineffective since they cannot handle legitimate demand paging [75, 59].

We use a database with 1,024 256 KB images from the Color FERET dataset [62], totaling 256 MB of sensitive data. We annotate the allocation of the database array to use ORAM (1 LOC) and compile with CoSMIX. As a result, the applica-
Table 5: Selective instrumentation of face verification server.

<table>
<thead>
<tr>
<th>Throughput (req/sec)</th>
<th>Native SGX</th>
<th>ORAM</th>
<th>ORAM → SUVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slowdown</td>
<td>8.6×</td>
<td>5.8×</td>
<td></td>
</tr>
</tbody>
</table>

We report the throughput achieved in Table 5. It shows that the ORAM mstore introduces an 8.6× slowdown. We note that for a dataset of 256 MB, the ORAM mstore overhead is about 41× more than native access, as reported in Table 3. This shows that selective instrumentation may make ORAM an attractive solution for some systems. However, with CoSMIX we can further reduce paging penalties.

**ORAM → SUVM stacking.** Although the application data is only 256 MB, PathORAM consumes about 860 MB due to its internal storage overheads. As a result, the SGX paging significantly affects the performance.

To optimize, we create a new ORAM → SUVM mstore by stacking ORAM on top of SUVM. Only 1 LOC in the ORAM mstore is annotated. The use of the combined ORAM → SUVM mstore improves the overall performance by 1.5× compared to the ORAM mstore alone (Table 5). Overall, selective instrumentation and mstore stacking result in a relatively low, 5.8× slowdown of the oblivious system compared to native SGX execution. This performance might be acceptable for practical ORAM applications and requires no code changes.

## 5 Related Work

**Enclave system support.** Recent works proposed systems to protect unmodified applications using enclaves [18, 21, 76, 80]. Other works proposed enclave enhancements [71, 47, 49], such as memory safety, ASLR, and enclave partitioning. Complementary to these works, CoSMIX provides system support for modular extensions to unmodified enclave applications.

**Trusted execution environments.** Previous works proposed different systems to protect applications from a malicious OS [30, 29, 42, 16, 15]. InkTag [42], for example, offers secured mmap service to applications; however, it relies on a trusted hypervisor with para-verification. CoSMIX puts its root of trust in hardware enclaves to implement in-enclave secure fault handlers.

**Controlled side-channel mitigation.** Previous works suggested the use of ORAM in SGX to improve its security [64, 66, 33, 86, 11, 50]. We believe that CoSMIX will allow the use of ORAM in many more applications via lightweight annotations. More efficient systems for mitigating the controlled side-channel attack have been proposed [22, 59, 75, 30]. For example, Apparition [30] uses Virtual Ghost [29], a compiler-based virtual machine, to restrict OS access to the page table. However, these systems do not support demand paging. CoSMIX protects applications that rely on demand paging from both a malicious OS and physical bus snooping attacks using the ORAM mstore.

**Customizing applications via paging mechanisms.** Previous systems proposed using page faults to improve performance and enable quality of service functionality in OS, GPUs, and enclaves [41, 60, 73, 14]. CoSMIX enables using similar enhancements in secure enclave systems.

Recent works take advantage of the RDMA infrastructure to enable efficient and transparent access to remote memory by customizing the OS page fault handler [39, 74, 53, 37]. For example, LegoOS [74] uses paging to simulate an ExCache for disaggregated memory support. CoSMIX takes a similar approach for extending secure enclaves.

Software-based distributed shared memory systems were proposed to customize memory access logic across remote machines [67, 68, 69, 63, 19, 55, 32]. These systems either use the page fault handler, runtime libraries or instrumentation of applications. CoSMIX is inspired by these systems; however, the mstore abstraction is more general and can support different memory access semantics, mixed or stacked in the same application.

**Memory instrumentation.** Much work has been done on instrumenting memory accesses using binary instrumentation tools [51, 24, 56, 52, 70] and compiler-based instrumentation [72, 47, 13, 31]. CoSMIX enables low-overhead selective memory instrumentation, specifically tailored for enclaves.

## 6 Conclusions

CoSMIX is a new system for modular extension of the memory layer in unmodified applications running in enclaves. It sidesteps the lack of hardware support for efficient and secure page fault handlers in existing architectures. CoSMIX enables low-overhead and selective instrumentation of memory accesses via a simple yet powerful mstore abstraction.

We show how compilation with CoSMIX both speeds up execution and adds protection to production applications. We believe that mstores may become a useful tool for facilitating the development of new secured systems.

## 7 Acknowledgments

We would like to thank our shepherd John Criswell for his valuable feedback. We also gratefully acknowledge the support of the Israel Science Foundation (grant No. 1027/18), the Israeli Innovation Authority Hiper Consortium, the Technion Hiroshi Fujiwara Cybersecurity center, and Intel ICRI-CI Institute grant, and the feedback from the Intel SGX SEM group. Meni Orenbach was partially supported by HPI-Technion Research School.
References


