StealthDB: a Scalable Encrypted Database with Full SQL Query Support

Alexey Gribov∗, Dhinakaran Vinayagamurthy†, Sergey Gorbunov†
∗Stealthmine Inc.
†University of Waterloo

Abstract—Encrypted database systems provide a great method for protecting sensitive data in untrusted infrastructures. These systems are built using either special-purpose cryptographic algorithms that support operations over encrypted data, or by leveraging trusted computing co-processors. Strong cryptographic algorithms usually result in high performance overheads (e.g., public-key encryptions, garbled circuits), while weaker algorithms (e.g., order-preserving encryption) result in large leakage profiles. On the other hand, some encrypted database systems (e.g., Cipherbase, TrustedDB) leverage non-standard trusted computing devices, and are designed to work around their specific architectural limitations.

In this work we build StealthDB – an encrypted database system from Intel SGX. Our system can run on any newer generation Intel CPU. StealthDB has a very small trusted computing base, scales to large datasets, requires no DBMS changes, and provides strong security guarantees at steady state and during query execution.

1. Introduction

Over the last decade, IT infrastructure has been undergoing major changes. Classically, enterprise data was held and processed within a company’s data center. Today, more and more companies are moving their data to third party public cloud infrastructure or service providers like AWS, Microsoft Azure and Google Cloud. These infrastructures have a few common features:

1) They are operated and maintained by potentially untrusted operators.
2) The infrastructure is shared between numerous clients. For instance, a single AWS physical instance may co-locate a number of virtual client instances.

Given these features, protecting confidentiality and integrity of user’s data from administrators, co-tenants, and other attackers is a major challenge.

To tackle this problem, research has been done to build “encryption-in-use” techniques that greatly improve security by preventing the attackers and even the cloud operators from ever seeing the data in clear. The state of art encryption-in-use database systems can be divided into two main categories: (A) systems built using advanced encryption schemes that allow to perform operations over the ciphertexts [PRZB11], [PSV+14], [PBP16], [PKV+14], and (B) systems that leverage a trusted processing device (e.g., FPGA) to perform operations [ABE+13], [BS11]. But, when attempting to provide reasonable performance overheads every encrypted database design inevitably leaks information about the underlying encrypted data, with some lower bounds known even for simply supporting encrypted search queries [GO96], [CT14]. So, all the systems in (A) had to balance between large leakage profiles, support for limited query functionality, or large performance overheads.

Allowing a trusted processing component helps the systems in (B) overcome some of the inherent limitations of systems in (A). But, each type of trusted component has its own set of limitations. So, the challenge in building systems in category (B) is in designing an architecture that addresses those limitations and simultaneously achieve the following1:

1. minimal changes to a DBMS engine,
2. small trusted computing base (TCB),
3. scale to large datasets with minimal performance overheads,
4. easy to deploy on a commodity system.

In this work, we study how to build an encrypted database system from a standard CPU leveraging Intel Software Guard Extensions (SGX) instruction set [MAB+13]. SGX is a small encrypted memory container (enclave) that can be accessed only by a predefined trusted code. The content of the enclave is protected from untrusted applications, system administrators, OS and hypervisor. SGX offers a great direction for protecting applications in cloud environments because of its strong security guarantees,

1. Some bullets are applicable for category (A) as well.
**SGX characteristic**  | **Design consequence**  
---|---
Small physical enclave size [MAB+13], [MAA+16], [AGJS13]  | Must keep amount of data minimal/constant at any given time.  
High overheads for random accesses inside enclaves [ATG+16]  | Design exitless communication between trusted/untrusted zones.  
High enclave exit/enter costs [OLMS17]  | Design communication interfaces between host kernel and enclave.  
No syscalls or IO [MAB+13], [MAA+16], [AGJS13]  | Must build data-oblivious code inside the enclave.  
SGX is vulnerable to side-channels [BMD+17], [LSG+17] [SWG+17], [WCP+17]  | Integrate application logic with SGX attestation services and application authentication mechanism for provisioning long-term secrets.  

| Attestation and key-provisioning [MAB+13], [MAA+16] [AGJS13]  
---|---

**TABLE 1: SGX architecture characteristics and design consequences.**

and general availability on all future release of Intel CPUs.

A few systems like [BS11], [ABE+13] were designed in a secure co-processor model, but it's not clear whether their design choices can be ported to the SGX architecture. SGX requires application rewriting to partition code into trusted and untrusted segments. Also, SGX is bounded to about 90 MB of processor reserved memory which is not nearly enough for even medium size database workloads. Databases need to scale to support arbitrary querying of gigabytes, or terabytes of data. Additionally, SGX is vulnerable to numerous side-channel leakages, lacks syscalls or IO support, and incurs high overheads for switching between enclave and non-enclave modes. In Table 1 we summarize some of the SGX limitations and consequences for any system leveraging it. As such, one cannot take a DBMS system and naively try to “run it in an enclave”.

Another challenge when building a database with SGX it to make sure that compromises in DBMS codebase or its authentication mechanisms do not result in data leakage. Also, database engines have been developed based on decades of research and development, and an ideal design must build on top of these results, without having to change their fundamental execution routines and principles.

### 1.1. Our Results

In this work, we ask the question of how to design an encrypted database system leveraging Intel SGX. Towards the goal, we first identify a desired list of security and functionality properties, along with a set of design goals and constrains. For instance, we believe that no or little changes to a DBMS are permissible. Also, data that is not actively queried should always be encrypted using a semantically secure encryption scheme. We refer the reader to Section 4.1 for a detailed list of our requirements and design constraints.

We investigate three possible design choices for an encrypted database with SGX (Section 4.2). Via a series of discussions and benchmarking experiments, we identify a design that works the best (Section 5). We implement this design on top of an existing postgres DBMS via a series of extensions and add-on modules. We call the resulting system StealthDB – a database engine where queries are processed in stealth mode. We show a high-level overview of our system in Figure 1. (Almost) no DBMS changes are required to use our system. Hence, any performance or feature improvements over the DBMS engines will directly translate to improvements in StealthDB. Our database scales to large datasets with a similar complexity to the native underlying DBMS engine, adding only a constant overhead for each query.

![Figure 1: High-level architecture overview of StealthDB](image)

We discuss various possible extensions in Section 5.5 and the security properties of our system in Sections 6. Our evaluation results show that the system can process OLTP queries with $\approx 1$ ms overhead over an unencrypted DBMS with $> 10$M overhead over an unencrypted DBMS with $> 10$M rows (or 2 GB plaintext) of a TPC-C warehouse database. We make conclusions and future directions for our system in Section 8.

2. Although various SGX extensions are promised by Intel in future releases, they are not available in the market yet and unclear when they will be. We also argue in the paper that these extensions should not affect our conclusions on the architecture of an encrypted DBMS with SGX.
1.2. Related Works

The work most similar to ours is Cipherbase [ABE+13]. In Cipherbase, computations over encrypted data are executed over deterministic (DTE) or order-preserving encryption (OPE), or offloaded to a trusted FPGA device. While our final design is similar to Cipherbase if all computations were offloaded to an FPGA, it was not clear prior to this work whether it would work best when FPGA is replaced with SGX enclaves. Our system is more secure than Cipherbase since we do not allow leaky DTE and OPE schemes. Moreover, FPGAs are not designed with security in mind for cloud-based applications. For instance, Cipherbase requires a trusted authority to load the secret key into the FPGA during a setup phase. Also, cloud providers usually implement their own layer of “shell” code inside the FPGA to control all the user code and I/O [Ama17]. It is not clear how to resolve these issues. On the other hand, though SGX comes with its own bag of limitations, we are able to design StealthDB around these.

TrustedDB [BS11] also uses a secure co-processor to perform operations. In their design, large portions of the DBMS engine (query parser and processor) are executed inside the TEE. We explore this design in Section 4.2 and conclude that it’s not ideal when working with SGX.

CryptDB [PRZB11] uses a hybrid of encryption schemes to support subset of SQL functionality. Their underlying large leakage profiles often result in data compromise [NKW15]. Mylar uses multi-key searchable encryption to protect web-applications [PSV+14]. Similar basic statistical attack can be applied on these encryption schemes [GMN+16]. Arx replaces OPE scheme with a special garbled-circuit-based searching method [PBP16]. Garbled circuits however introduce large computational and storage overheads.

A few works studied how to build versions of encrypted databases with SGX. Opaque studies how to leverage SGX to secure distributed analytical workloads in Spark systems [ZDB+17]. VC3 system proposes and architecture for analytical MapReduce jobs in cloud settings [SCF+15].

HardIDX investigates how to perform index searches over BTrees in an enclave [FBB+17]. They consider two design choices: first design in which the entire BTree is loaded into an enclave, decrypted and processed in cleartext, and the second design where parts of BTree are loaded during query processing. Their conclusions are similar to ours (Section 4.2), where we show that databases operating over large datasets scale better when the amount of off code and data in an enclave is kelp small. Overall, [FBB+17] just prototypes index searches, whereas we architecture a complete encrypted database system.

A number of works study how to load unmodified applications into enclaves [BPH14], [ATG+16], [TAB+14], [HZX+16]. These approaches work well for applications that process small data sizes, but do not scale well to larger workloads due to SGX limitations. Also, increasing the complexity of the codebase inside the enclaves aggravates the security risks associated with SGX [LJJ+17].

OSPIR-OXT [CJJ+13], [CJI+14], [FIK+15], SisoSPIR [IKLO16] and BLIND SEER [PKV+14] build encrypted database systems from scratch with provable security guarantees for a subset of functionality based on different cryptography tools. There are also multitudes of other works which provide improvements over security or specific functionalities of a database, but they are not implemented or integrable with an existing database. A recent SoK paper provides a great summary of the state-of-art research in encrypted database systems [FVY+17].

Fully homomorphic encryption [Gen09] is another powerful cryptographic primitive which enables an untrusted user to perform arbitrary computations on encrypted data without learning any information about the underlying data. But the current constructs for doing this are very far from being practical [HS14]. In general, while theoretical security of systems built based on cryptographic methods can be high, the real-world security of the system relies on the multitude of factors: correct implementations of non-trivial crypto algorithms, meta-data contents, DBMS structure and stored relationships in data-structures, information in log files, etc.

We summarize the comparison with the related encrypted database systems in Table 2.

2. Background on Intel SGX

In this section we give a brief introduction to Intel Software Guard Extensions (SGX). We refer the reader to [MAB+13], [CD16] for more details on SGX. Intel SGX is a set of new x86 instructions that enable code isolation within virtual containers called enclaves. In the SGX architecture, developers are responsible for partitioning the application into enclave code and untrusted code, and to define an appropriate I/O communications interface between them. In SGX, security is bootstrapped from an underlying trusted processor, but not trust in a remote software stack. On the high level, to a user the SGX hardware presents the Load(P) and Execute(Ep,input) functionalities.

\[ \text{Load}(P) \rightarrow (E_P, \phi) \\] The load function creates an enclave with an identifier \( E_p \) and loads the program

\[ \text{Execute}(P, \phi) \rightarrow \phi \\] The execute function runs the program and returns the result.
TABLE 2: Summary of security, supported functionality, performance overhead and the ease of integration with the underlying database for the existing encrypted relational databases and ours. Most of the entries are taken or inferred from [FVY+17] to the best of our knowledge of the practical attacks.

<table>
<thead>
<tr>
<th>System</th>
<th>Security assumptions</th>
<th>Leakage</th>
<th>Integrity</th>
<th>Functionality</th>
<th>Engine type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>steady state</td>
<td>runtime</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CryptDB [PRZB11]</td>
<td>Crypto</td>
<td>●</td>
<td>●</td>
<td>×</td>
<td>●</td>
</tr>
<tr>
<td>Cipherbase [ABE+13]</td>
<td>FPGA+Crypto</td>
<td>○</td>
<td>○</td>
<td>×</td>
<td>○</td>
</tr>
<tr>
<td>TrustedDB [BS11]</td>
<td>IBM SCpus</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>○</td>
</tr>
<tr>
<td>Arx [PBP16]</td>
<td>Crypto</td>
<td>○</td>
<td>○</td>
<td>×</td>
<td>●</td>
</tr>
<tr>
<td>BLIND-SEER [PKV+14]</td>
<td>Crypto</td>
<td>○</td>
<td>○</td>
<td>✓</td>
<td>●</td>
</tr>
<tr>
<td>OSPIR-OXT [FJK+15]</td>
<td>Crypto</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td>StealthDB</td>
<td>SGX</td>
<td>○</td>
<td>○</td>
<td>×</td>
<td>●</td>
</tr>
</tbody>
</table>

- Steady state is the hypothetical state of the system when no queries are being processed, for a system which had been processing queries before entering this state. We consider this because some systems leak more information after executing a few queries than during initialization.
- Runtime refers to the state of the system when processing a query.
- Partial leakage can be of different types: inequalities between encrypted values query, statistics of query output, access patterns in memory and disk etc. Our estimates in the leakage columns are relative as in [FVY+17].
- Semantic security of the database does not prevent leakage of the structure of the database like the number of rows and columns in a table and so on.
- Query engine type specifies the parts of the underlying plaintext DBMS that is (or needs to be rebuilt) to use the system. Custom refers to a complete rebuild, and legacy refers to using the underlying system as such.

A client receives a proof $\phi$ that its intended program $P$ (and initial data) has been loaded into an enclave. The proof $\phi$ can be used by the client to attest that the right program has been loaded inside an enclave with respect to a measurement (hash) of that program.

$\text{Execute}(E_P, \text{input}) \rightarrow (\text{out}, \psi)$. The execute function is given an enclave $E_P$ handle (corresponding to an enclave with a program $P$), it then runs it on an input input, to produce a tuple constituting of the output out and a proof $\psi$ which the client can use to verify that the output out was produced by the enclave $E_P$ executing with input.

There are three main functionalities that enclaves achieve: isolation, sealing and attestation.

**Isolation:** code and data inside the enclave protected memory cannot be read/modified by any process external to the enclave. SGX does this by isolating enclave code and data in the Processor Reserved Memory (PRM), referred to as Enclave Page Cache (EPC), which is a subset of DRAM that gets set aside securely at boot time. Cache lines read into the processor cache from the EPC are isolated from non-enclave read/writes via hardware paging mechanisms, and encrypted/integrity checked at the processor boundary. Cryptographic keys for these operations are owned by the trusted processor. Thus, data in the EPC is protected (privacy and integrity-wise) against certain physical attacks (e.g., bus snooping), the operating system (direct inspection of pages, DMA), and the hypervisor.

**Sealing:** data passed to the host environment is encrypted and authenticated with a hardware-resident key. Every SGX processor has a Root Seal Key that is embedded during the manufacturing process. An enclave can derive a Seal Key that is specific to the enclave identity from the Root Seal Key and this Seal Key can be used to encrypt/authenticate data and store it in untrusted memory. Sealed data can be recovered by the same enclave even after enclave is destroyed and restarted on the same platform. But the Seal key cannot be derived by a different enclave even after enclave is destroyed and restarted on the same platform. We will use the following Seal and Unseal algorithms:

$\text{Seal}(\text{AAD}, \text{msg}) \rightarrow \text{seal\_ct}$
$\text{Unseal}(\text{AAD}, \text{seal\_ct}) \rightarrow \text{msg} / \perp$

SGX uses AES-GCM to encrypt $\text{msg}$ using the derived Seal key. Here, $\text{AAD}$ is the additional authentication data which is included as a part of the MAC to provide integrity but not encrypted along with $\text{msg}$. We will ignore the AAD argument when there is none.
**Attestation:** a special signing key and instructions are used to provide an unforgeable report attesting to code, static data, and (hardware-specific) metadata of an enclave, as well as outputs of computations performed inside the enclave. There are two forms of attestation: local and remote.

- **Local attestation.** An enclave $A$ uses local attestation to attest to another enclave $B$ on the same platform. Since enclaves on the same machine share the same Root Seal Key, the enclave $A$ uses a special instruction which creates a MAC of its measurement and its metadata (along with additional optional data provided as input to the instruction) with a *Report Key* corresponding to the enclave $B$ deriving from the Root Seal Key. The resulting MAC is called a *report*. Now, the enclave $B$ can verify the report by deriving the same Report Key from the Root Seal Key.

- **Remote attestation.** Remote attestation generates a report that can be verified by any remote party. Roughly, an enclave first local attests to a special enclave called the Quoting Enclave (QE), sending it a report. The QE verifies local reports and if valid, signs the same underlying data with a private key for an anonymous group signature scheme called Intel Enhanced Privacy ID (EPID) [JSR+16]. The QE obtains this private key during initialization with the Intel Provisioning Server upon device initialization. The resulting signature is called a *quote*. Currently, the remote party requires contacting the Intel Attestation Server to verify quotes, though in principle this could be done by any verifier that has the group public key.

**Key establishment during attestation.** Key establishment between two enclaves or between an enclave and a remote party can be accomplished on top of the local/remote attestation process. An enclave can send the key shares (for e.g., a Diffie-Hellman key share $g^a$) and include them as the *additional authentication data* to MAC. Thus attestation provides authenticity and integrity to the key share from the enclave. In our system, we will very often run the key establishment phase on top of local/remote attestation to establish a secure channel for communication between two enclaves or between an enclave and a remote party using the established shared secret key. We will use the following two pairs of function calls to achieve these tasks:

- \[
    \text{LocalAttest} + KE_{\text{src}}(\text{dest_enclave}) \rightarrow k / \perp
    \]
- \[
    \text{LocalAttest} + KE_{\text{dest}}(\text{src_enclave}) \rightarrow k / \perp
    \]
- \[
    \text{RemoteAttest} + KE_{\text{src}}(\text{dest_enclave}) \rightarrow k / \perp
    \]
- \[
    \text{RemoteAttest} + KE_{\text{dest}}(\text{src}) \rightarrow k / \perp
    \]

Here, $k$ is the key established between the source and destination enclaves if the attestation completes successfully, and $k$ will be used to encrypt the further communication between them. (During remote attestation, the src need not be an enclave.)

**SGX TCB.** SGX stands out in that its TCB consists only of the CPU microcode and privileged containers, however it also requires the user to trust in Intel’s key management infrastructure for signing microcode and various service enclaves. In particular, we must trust that the root seal keys embedded into devices are not leaked from the manufacturing facility, and that the Intel Provisioning Server safely manages root provisioning keys as well as EPID master secret keys.

**Design challenges with Intel SGX.** In Table 1 we summarize the properties of Intel SGX (1.0) that make designing a system based on SGX challenging in terms of both security and performance. Currently, the size of EPC is physically upper bounded by 128 MB by the processor. Around 30 MB of EPC is used for bookkeeping, leaving around 95 MB of usable memory. To support applications with large working sets, the OS performs paging to move pages in and out of the EPC on demand. Hardware mechanisms in SGX ensure that all pages swapped in/out of the EPC are integrity checked and encrypted before being handed to the OS. Thus, the OS learns only that a page with a public address needed to be swapped, not the data in the page. Special pages controlled by SGX (called VA pages) implement an integrity tree over swapped pages. In the event the system is shutdown, the VA pages and (consequently) enclave data pages are lost. However EPC paging is expensive and can cost between 3x and 1000x depending on the underlying page access pattern (Figure 3 in [ATG+16]). At any time, the OS controls when enclave code starts and stops running. Each switch incurs a large performance overhead – the processor must save the state needed to resume execution and clear registers to prevent information leakages. Exitless communication mechanisms avoid unnecessarily context switches via a queue shared between enclave and untrusted application logic [OLMS17].

Although SGX prevents an adversary from directly inspecting/tampering with the contents of the EPC, it does not protect against multiple software-based side channels. Correspondingly, the literature has demonstrated attacks that extract sensitive data through hardware resource pressure (e.g., cache [BMD+17], [SWG+17], thread scheduling [WKP16] and branch predictor [LSG+17]) and the application’s page-level access pattern [XCP15]. Many of these works also provide fixes for their attacks with varying overheads.
3. Platform Overview

3.1. Usage Model.

We work with the following setting. A data owner wishes to store data securely on a remote untrusted SQL database server. He also wants to support authorized clients to issue queries. The data owner authorizes clients by issuing them credentials. The server will then maintain a credential database for the authorized clients (in an encrypted form). Each user authenticates to the server using its credentials, which allows her/him to issue certain types of queries on the database. The server in our model is equipped with a secure processor, such as Intel SGX. Hence, the server can be identified with some “platform-key” established by Intel SGX. This key will be used during the attestation protocols. The data owner and clients engage in attestation and transfer any secret materials (master key, credentials, etc.) to the SGX enclaves via secure channels.

3.2. Threat Model

StealthDB provides security against passive adversaries. A passive adversary does not inject malicious code or alter the program execution in any way, but it can read the contents of the memory, disk and all the communication, and hence may passively attempt to learn additional information from the data they observe.

There are two dimensions in which we analyze the threat model for our system. The first dimension is about adversaries restricted to accessing only the disk versus the adversaries being able to access both the memory and disk used by the cloud provider. The second dimension is about whether the adversaries can get snapshot accesses to memory and disk versus those that get persistent access.

StealthDB satisfies our security goals for each adversarial type. We will provide a detailed discussion of the leakage profile of our system for each adversarial type and various attack vectors in Section 6.

4. Designing an Encrypted Database

In this section, we describe a few design goals we set out to achieve for our system, and discuss and experiment with a few possible design choices that one needs to consider when building an encrypted database from SGX.

4.1. Design Goals

There is a three way trade-off between security, functionality and performance while designing an encrypted database. Often, an optimal trade-off depends on the underlying data and query types that the DBMS needs to support. In this work, we focus on building a scalable encrypted database system that can support arbitrary query types, with minimal possible leakage. We outline some of the design goals for StealthDB:

- The performance goal is to achieve a DBMS query runtime that scales identically to the native DBMS engines. That is, any query that takes $T$ time to execute in a native DBMS, should take at most $c \cdot T$ time to execute over encrypted data, for some constant $c$.
- The functionality goal is to support the full SQL functionality of the native DBMS.
- At a high level, the security goal is to provide semantic security of data and index structure on disk at all times and semantic security of cold in-memory cached data (i.e. when queries are not executed). But, we do end up leaking the data relationship structure that is produced by a DBMS engine during query execution in memory. This structure may contain information about individual encrypted data values like the inequalities between them. This leakage, however, seem inherent to achieve the desired efficiency goal. We refer the reader to Section 6 for more details.

4.2. Designing an Encrypted DBMS from SGX

We consider three design choices and evaluate them on a few micro experiments to help us understand how to build an encrypted database system with SGX. The design choices are summarized in Figure 2. We envision that in all three design choices data is encrypted on disk using a semantically secure encryption scheme. The designs differ in how queries are executed over the data.

The first, most obvious design would be to run the entire DBMS inside an enclave (left figure in 2). The data would be read from disk, decrypted transparently and then the DBMS would perform all necessarily operations inside an enclave. By default, SGX, however, is not well suited for this task for a few reasons. First of all, it does not have IO or syscall support, so an additional outside shim layer would need to be exposed to talk to the kernel level, and application dependencies need to be loaded inside (or outside via shim) an enclave. We consider this as moderately hard, since a few academic projects such as Scone [ATG+16] and Graphene [TAB+14], [cTPV17] show how to load unmodified executables into enclaves. Also, SGX is currently limited to 90 MB of working memory and significant penalties are paid for going beyond that limit [OLMS17]. Future releases of SGX promise larger enclave sizes. However, inherent Merkle
tree integrity protection on each page, to prevent replay attacks, does not scale well asymptotically to larger enclaves. Moreover, this design would keep a very large TCB inside the enclave: the entire DBMS engine, any communication logic with the “outside world” and dependencies. Finally, SGX is vulnerable to numerous side-channels and hence very custom modifications to the DBMS would still need to be performed to prevent these attacks and make code oblivious.

The second design we consider (middle figure in 2) keeps most of the DBMS in the untrusted zone. However, it places the query execution logic in the enclave. That is, when a query needs to be executed, individual tables can be brought into the enclave to perform selections, projections, joins, etc. The query plan, I/O and other DBMS parts remain in the untrusted memory. In terms of scalability, this design suffers from the same problems as the previous choice. Also, tables/indexes need to read from disk, deserialized and then loaded into enclave. In Figure 3 we show that performance overhead for this operation over native DBMS is around $3 \times$ when the dataset fits within an enclave, and goes up to $9 \times$ for large datasets. Similarly, query processing logic would need to address SGX side-channels. Finally, partitioning a DBMS into this architecture is a challenging task.

In the last design, we keep most of the DBMS in the untrusted zone. At the lowest level of the parsed query tree, each query is eventually broken down into some primitive operators (e.g., $<=$, $>$, $+$, $*$) over individual data values. To perform operations over encrypted data in this design, we transfer individual data item(s) to an enclave, followed by decryption, operator function and encryption inside the enclave. Hence, the data relation structure produced by the query is kept in the untrusted memory. The advantage of this design is that the communication with the disk and network layers would remain unchanged. Overall, minimal changes to the DBMS are needed – one only needs to change how primitive operators on data values are performed (see Section 5). Also, the amount of code/data inside an enclave can remain very small (constant), since the entire datasets would reside in the untrusted memory. This keeps the TCB very small, and it is easy to make it data-oblivious. However, one leaks relationship between encrypted data values during query execution in this design.

In Figure 4, we compare performance of performing B-tree searches over database indexes in later two design choices. As expected, one can see that performing a search when an entire B-tree is loaded inside an enclave does not scale to larger datasets. (However, it performs well when the tree size is very small and can be fit entirely into an enclave.) In the third design, when the B-tree is kept encrypted in the untrusted memory but individual comparisons are executed in an enclave, we see up to $100 \times$ overheads compared to performing the search over unencrypted data. This is explainable by high switching costs of ocall/ecall functions. Using an exit-less communication mechanism via a shared queue [OLMS17], we can reduce this overhead by $5 \times -10 \times$.

5. Architecture

The architecture of StealthDB is presented in Figure 5. StealthDB makes extremely minimal changes to the underlying DBMS, with most of our components augmented on top of an unmodified DBMS.
Figure 4: Latency to execute random binary tree searches comparing different approaches. Two different implementations of the partial approach: comparison function as trusted calls and the exit-less communication via a queue for transferring data to/from an enclave.

We will now go through the flow of a database creation, query life-cycle, and explain each of our components in details below.

5.1. Database creation

When a database is created, the database owner designs a database schema to define the structure of the database. During the schema creation, StealthDB allows the owner to identify the columns of the tables in the database which have sensitive information and use our encrypted datatypes for those columns. An encrypted datatype is used to represent values which are the encrypted versions of its corresponding plaintext datatype. For instance, encrypted integers are represented by the encrypted datatype $\text{encint}$.

![Figure 6: Definition of $\text{encint}$](image)

A database owner may issue the following command to create a table with two columns of types encrypted integers and encrypted strings:

```sql
CREATE TABLE item (
  id encint NOT NULL,
  name enccstring NOT NULL,
);
```

StealthDB will encrypt the data values in an encrypted datatype using AES-GCM which is an authenticated encryption scheme providing both confidentiality and integrity of the data values. We will discuss about the key(s) used by this encryption during the DBMS initialization.

5.2. DBMS Initialization

When the DBMS is started, the following additional steps are performed for StealthDB.

**Enclaves creation.** StealthDB creates three enclaves on the database server: the client authentication enclave Auth, the query pre-processing enclave PreProcessor and the operation enclave Ops. The full descriptions of the programs that are to be loaded inside these enclaves is described in Figures 8, 9, and 11, respectively. These enclaves are loaded by an untrusted DBMS runtime, but our system will later allow to attest that the correct enclaves have been initialized. That is, to ensure the correctness of the loaded programs in the enclaves, the clients use the remote attestation process and the publicly available measurements (hash) of the enclave code. We will defer the explanation of this step and the functionality of these enclaves to the sections below.

To facilitate communications between users and enclaves, StealthDB introduces an I/O layer on the server side. Its job is to simply redirect requests between the appropriate enclaves and the DBMS. This will also act as the wrapper program for the enclaves helping in processing their I/O requests and system calls. Note that this layer is untrusted and can be controlled by an adversary.

**Key generation.** The initialization phase also involves generating a master secret key. StealthDB performs key generation inside the Auth enclave. Auth runs the KeyGen() function to sample a 128 bit secret key $K$ at random for the AES encryption/decryption operations. In the current design, this master key $K$ will be used to encrypt all the data values in the database. (We do this for simplicity and our design can be extended to support an integration with a key management service to enable the usage of different keys for different clients or for different columns in the database).

The master key $K$ is then transferred to the PreProcessor and the Ops enclaves as follows. When the PreProcessor and Ops enclaves are created, they perform a local attestation with Auth as described in Section 2 and establish a secure channel with Auth. When the attestations succeed and after the secure channels are established, Auth’s KeyTransfer() function uses the channels to send the master key.
K to PreProcessor and Ops. (On the other end, PreProcessor and Ops will run their KeyReceive() function to complete these steps and receive K). On obtaining K, PreProcessor and Ops use SGX’s sealing property to encrypt and store K for future use.

**Transfer of credentials.** The final task of the initialization phase involves transferring the client credentials and access policies to Auth. A client (proxy) will authenticate to Auth. And, from the point of view of the DBMS, Auth (and PreProcessor) will act as a client who has complete access to the database. To facilitate this, the data owner first engages in a remote attestation protocol with Auth along with a secure channel establishment and if it succeeds, he/she sends the master credentials along with the database of client credentials and access policies to Auth through the established channel. On obtaining these, Auth uses the SGX seal operation to encrypt and store them.

### 5.3. Client authentication

One of the challenges we need to address is to make sure only authorized users can query the encrypted database system. For this, we design an authentication method built on top of an existing DBMS.

After the database server is started, it is now ready to accept connections from the clients. Here, StealthDB augments an authentication between the client and the Auth enclave such that clients can authenticate to the Auth enclave. This works as follows.

First, the client proxy verifies that the DBMS has loaded the correct code into Auth, by performing the remote attestation (plus secure channel establishment) protocol with Auth as described in Section 2. Let sessk be the shared secret key obtained after its successful completion. The client will then authenticate to the Auth enclave using its credentials, say its password or its SSH key, through the established secure channel. On the server side, the I/O layer directs the client...
**Auth enclave**

- **CompleteAttest(src_enclave)**
  1) Complete the local attestation initiated by src_enclave to obtain a shared secret key for Auth to communicate with src_enclave: LocalAttest + KE_{dest}(src_enclave) → \( k_{Auth,src_{enclave}} \)
  2) Run \( \text{Seal}(k_{Auth,src_{enclave}}) \) → seal_{ctAuth,src_{enclave}} and store the sealed key as seal_{ctAuth,src_{enclave}}.

- **KeyGen()**
  1) Choose \( K \sim \{0,1\}^{128} \) and let \( K \) be the master secret key.

- **KeyTransfer(dest_enclave)**
  1) Check if seal_{ctAuth,dest_{enclave}} exists. If not, call the \( \text{Attest(Auth)} \) function in dest_enclave.
  2) Unseal seal_{ctAuth,dest_{enclave}} to get the shared secret key: \( \text{Unseal}(\text{seal}_{ctAuth,dest_{enclave}}) \rightarrow k_{Auth,dest_{enclave}} \)
  3) Encrypt the master secret key as \( \text{AES.Encrypt}_{Auth,dest_{enclave}}(K) \rightarrow ct \).
  4) Send ct to dest_enclave.

- **CompleteClientAuth()**
  1) Complete the remote attestation initiated by the client proxy RemoteAttest + KE_{dest}(client proxy) → sessk and obtain the session key sessk shared with the client.
  2) Validate the client credentials using the credential database.
  3) If it succeeds, run TokenTransfer(ID, sessk) to send the client ID and the session key to the PreProcessor enclave.

- **TokenTransfer(ID, sessk)**
  1) Check if seal_{ctAuth,PreProcessor} exists. If not, call the \( \text{Attest(Auth)} \) function in PreProcessor.
  2) Unseal seal_{ctAuth,PreProcessor} to get the shared secret key: \( \text{Unseal}(\text{seal}_{ctAuth,PreProcessor}) \rightarrow k_{Auth,PreProcessor} \)
  3) Encrypt the client ID and session key as \( \text{AES.Encrypt}_{PreProcessor}(\text{ID}|\text{sessk}) \rightarrow ct_{sess} \).
  4) Send ct_{sess} to dest_enclave.

**PreProcessor enclave**

- **Attest(dest_enclave)**
  1) Run LocalAttest + KE_{dest}(dest_enclave) → \( k_{dest_{enclave},PreProcessor} \) to obtain a shared secret key for PreProcessor with dest_enclave.
  2) Run \( \text{Seal}(k_{dest_{enclave},PreProcessor}) \) → seal_{ctdest_enclave,PreProcessor} and store the sealed key as seal_{ctdest_enclave,PreProcessor}.

- **KeyReceive(ct)**
  1) Unseal seal_{ctAuth,PreProcessor} to get the shared secret key with Auth: \( \text{Unseal}(\text{seal}_{ctAuth,PreProcessor}) \rightarrow k_{Auth,PreProcessor} \)
  2) Decrypt the input ciphertext to obtain the master secret key:
     \( \text{AES.Decrypt}_{k_{Auth,PreProcessor}}(ct) \rightarrow K \).
  3) Run \( \text{Seal}(K) \rightarrow \text{seal}_{ctMaster} \) and store the sealed key as seal_{ctMaster}.

- **TokenReceive(ct_{sess})**
  1) Unseal seal_{ctAuth,PreProcessor} to get the shared secret key with Auth:
     \( \text{Unseal}(\text{seal}_{ctAuth,PreProcessor}) \rightarrow k_{Auth,PreProcessor} \)
  2) Decrypt the input ciphertext to obtain the client identity and the corresponding session key:
     \( \text{AES.Decrypt}_{k_{Auth,PreProcessor}}(ct_{sess}) \rightarrow (\text{ID}|\text{sessk}) \).
  3) Run \( \text{Seal}(\text{ID},\text{sessk}) \rightarrow \text{seal}_{ctID} \) with ID as the additional authentication data and store the pair (ID, seal_{ctID}).

- **QueryPreProcessing(ID, ct_query)**
  1) Unseal seal_{ctID} to get the session key sessk for the client ID: \( \text{Unseal}(\text{seal}_{ctID}) \rightarrow \text{sessk} \)
  2) Decrypt \( ct_{query} \) using sessk to obtain the query: \( \text{AES.Decrypt}_{sessk}(ct_{query}) \rightarrow \text{query} \).
  3) Check if the client is permitted to run this query: \( \text{QueryControl}(\text{ID}, \text{query}) \rightarrow 0/1 \). If 0, abort.
  4) Unseal seal_{ctMaster} to get the master secret K: \( \text{Unseal}(\text{seal}_{ctMaster}) \rightarrow K \)
  5) Parse query and encrypt the data values in it using the master secret K: \( \text{QueryParser}(K, \text{query}) \rightarrow \text{encquery} \)
  6) Output encquery

Figure 9: Pre-processing enclave is used to prepare queries arriving from a user for DBMS execution.

The session key for the client.

Once the client authentication is completed, the interaction with the client for query processing will be performed by the PreProcessor enclave. To facilitate this, the I/O layer will now invoke the TokenTransfer(ID, sessk) function in Auth to transfer...
the client “ID” and sessk to PreProcessor. This transfer will use the secure channel established between these enclaves during the master key transfer. The TokenReceive function of PreProcessor will seal and store sessk with ID as the additional authentication data during the seal operation.

5.4. Query execution

Now we will explain the working of query processing and execution in StealthDB for a client who has completed its authentication successfully. The design of StealthDB permits the use of an unmodified query driver (e.g., JDBC, ODBC, etc.).

When a client issues a query, the client proxy encrypts the entire query string using the session key sessk with its ID part of the authenticated values. On the server side, the I/O layer directs the client queries to PreProcessor. The QueryPreProcessing function first decrypts the query ciphertext using the session key sessk for ID. Then, it checks whether this client is permitted to run this query. Typically, a DBMS allows the DB owners to specify access control policies for the clients. Here, we rewrite the access control monitor inside PreProcessor and the check can be invoked with the QueryControl function. If the checks are passed, QueryPreProcessing runs our version of QueryParser with the client query query as input. QueryParser identifies the data values in the query which correspond to the columns in the database using encrypted datatypes, and AES encrypts these data values using the master secret key K.

The output encquery of this step is given to the DBMS for execution. Note that the DBMS is oblivious to the changes made to the query. The structure of encquery is same as that of the query issued by the client. This lets the DBMS use an unmodified query parser to parse this query. But after the query is parsed and a query plan is obtained, we need to augment the DBMS with functions to operate on the encrypted datatypes. We do this as follows.

We first identify the set of primitive operators used by the underlying DBMS. Primitive operators are those further-indivisible operators used in query plans:

- **Comparators** such as <, >, <=, >=, !=, etc.
- **Math operators** such as +, -, %, *, etc.
- **Hash functions** that are used to build some indexes.
- **Advanced math functions** such as sin, cos, tan, etc.

Traditionally, DBMSs define a functionality for each input datatype tuple supported by a primitive operator. StealthDB augments these with their functionalities when used with the corresponding encrypted datatypes as in Figure 10.

![Figure 10: Operator = for encint. Here, encint_eq will call the Ops enclave to decrypt the input, check their equality and output the result.](image)

Our implementation on Postgres implements primitive operator functionalities over the encrypted datatypes and include them as extensions.

For every possible input datatype tuple, we define a function inside the Ops enclave. Suppose that we are given two encrypted data values (e1, e2) and an operator ⊕, the corresponding function inside Ops will perform:

1) decryption of the inputs e1, e2 using the master key to get plaintext values p1, p2,
2) perform the operator function to get p\* = p1 ⊕ p2,
3) encrypt the result p\* to get a ciphertext e\* using the master key (if specified by the design).

The number of inputs and outputs may of course vary depending on operator. Moreover, datatype conversion are also allowed in the model. For example, an encrypted integer may be converted to an encrypted string, and so on. Thus, we only perform a few basic operations (decrypt, primitive operator, encrypt) during the query execution inside the enclave.

Standard SGX ocall/ecall communication mechanism with enclaves is too slow when many calls are needed. To solve this, we implement an exit-less mechanism [OLMS17] for communicating with Ops. In [OLMS17], there is always one thread running inside an enclave listening for operator jobs. The DBMS uses our I/O layer to send jobs and receive replies via a communication queue. This method greatly improves performance by avoiding context switch for each call to the operator between trusted and untrusted zones. Please refer to Section 4.2 for a detailed discussion on this.

There are also other inherent advantages with our design:

- When a client issues a query only involving unencrypted datatypes, the query processing and execution proceeds in the native way and hence with no overheads.
- A very interesting property is that our design also allows for computations between encrypted datatypes and unencrypted datatypes. The database
Ops enclave

- **Attest(dest_enclave)**
  1) Run LocalAttest + KESEK(dest_enclave) → $k_{\text{dest enclave,Ops}}$ to obtain a shared secret key for Ops with dest_enclave.
  2) Run Seal($k_{\text{dest enclave,Ops}}$) → $\text{seal}_ct_{\text{dest enclave,Ops}}$ and store the sealed key as $\text{seal}_ct_{\text{dest enclave,Ops}}$.
- **KeyReceive(ct)**
  1) Unseal $\text{seal}_ct_{\text{Auth,Ops}}$ to get the shared secret key with Auth: Unseal($\text{seal}_ct_{\text{Auth,Ops}}$) → $k_{\text{Auth,Ops}}$
  2) Decrypt the input ciphertext to obtain the master secret key: AES.Decrypt($k_{\text{Auth,Ops}}(ct)$) → $K$.
  3) Run Seal($K$) → $\text{seal}_ct_{\text{master}}$ and store the sealed key as $\text{seal}_ct_{\text{master}}$.
- **encrypted_primitive_operator({enc_input})**
  1) Unseal $\text{seal}_ct_{\text{master}}$ to get the master secret $K$: Unseal($\text{seal}_ct_{\text{master}}$) → $K$
  2) Decrypt each enc_input using $K$: sgx_rijndael128GCM_decrypt($K$, enc_input) → input
  3) Perform the computation as specified by the operator to obtain result.
  4) If result needs to be encrypted, run sgx_rijndael128GCM_encrypt($K$, result).

Figure 11: Operations enclave is used to perform basic operations over the encrypted data values.

- The owner here can also specify that the output of such computations should be encrypted to avoid leaking information about the encrypted inputs.
- Since our design implements only the primitive operators, it is easy for us to implement them inside Ops using data-oblivious methods [OSF+16] with a very small performance overhead to counter the side-channel attacks of SGX.

5.5. Extensions

*Encrypting indexes* The indexed columns, unlike the other columns in the table, need an extra layer of protection. The database owner can indeed specify in the database schema if the data values in the indexed column need to be encrypted. But, when the column is indexed into a B-tree, for example, the structure of the tree reveals the inequalities with respect to the values in the column even though the individual values in the tree are encrypted. The inequalities are available even to a passive adversary after index creation before any query is made to the database.

To avoid this leakage on disk, StealthDB encrypts every page that is written to the files on disk corresponding to the indexes. We do this by making minimal modifications to the underlying DBMS by encrypting the data right before it is written to the index files on disk, and decrypting the data read from the index files right after it is read from disk. In our implementation for Postgres, our changes to the codebase involve adding three lines of code to do this task. We create and run a fourth enclave $\text{Index}_{\text{OP}}$ during the DBMS initialization which performs the encryption and decryption of the index data pages. And the three new lines are for retrieving the enclave ID, calling the encryption function inside $\text{Index}_{\text{OP}}$ right before a $\text{FileWrite}$() of Postgres and for calling the decryption function inside $\text{Index}_{\text{OP}}$ right after a $\text{FileRead}$(). The key used for these routines is generated and stored by $\text{Index}_{\text{OP}}$, and Auth attests the correct loading of $\text{Index}_{\text{OP}}$ during the DBMS initialization.

*Encrypting logs.* Some of the log files reveal sensitive information about the queries even for a snapshot adversary on disk [GRS17]. StealthDB protects against an adversary accessing disk by encrypting the log files on disk in a way similar to our encryption of index files on disk.

*Key management.* In the current implementation, we use a single master key $K$ to encrypt all the data values. $K$ is sealed and stored on the disk by $\text{PreProcessor}$ or $\text{Ops enclave}$ when obtained from Auth. As explained in Section 2, if and when the system is restarted, the enclaves are created again and a valid $\text{PreProcessor}$ or $\text{Ops enclave}$ can unseal the corresponding sealed components to obtain $K$. During this process, the AES-GCM encryption used in the SGX sealing provides confidentiality and integrity for the sealed component of $K$ against any adversary.

6. Security

We will first detail the leakage profile of StealthDB and through a series of security claims we will argue that StealthDB does not leak any more information than what is part of the leakage profile. Our evaluation is with respect to the architecture we propose, and hence independent on the specific underlying DBMS engine.

6.1. Leakage profile

StealthDB attains the following leakage profile as a result of the security vs functionality vs performance tradeoff decided by our design.
Initialization.
- During the initialization protocol where the data owner uploads a database to the server, StealthDB provides semantic security to all the data stored by the DBMS.

Query phase. When a query is being executed,
- a passive adversary with only a snapshot access to the memory does not learn the individual data values in the index (since they are encrypted with semantically secure encryption), but it learns the structure of the index (for e.g., the structure of B-tree for a B-tree indexed column).
- a passive adversary with persistent access to the memory can learn information revealed by the access patterns leakages for the query being executed.

An example of an access pattern leakage in memory is as follows. When a comparison is made between two encrypted data values of an indexed column, the resulting branch of the B-tree being accessed reveals the result of the comparison. For instance, given two encrypted ciphertexts $ct_1, ct_2$ corresponding to plaintext integers $a, b$, the DBMS server should be able to learn whether $a \leq b$ (if requested by the user query), but nothing else. The memory trace of a persistent adversary will include $ct_1, ct_2$ and the result of the inequality relationship. We emphasize that the result of the inequality relationship is the minimum leakage that we believe is necessary to achieve the standard DBMS performance goals. An example of an access pattern leakage in disk is as follows. During an insert query, the index pages on disk that are being modified reveal some inequalities with the data being inserted.

The precise information revealed by query access patterns depends on the underlying DBMS. But we believe that a practical DBMS would attempt to minimize the data being accessed to process a query, and hence a passive adversary is restricted to learning access pattern leakages only for the data relevant to the query. The data irrelevant to the query will remain semantically secure.

Finally, StealthDB also provides semantic security to the whole database during a hypothetical steady state. That is, even when the DBMS is up and running, when no queries are being executed at an instant and when the changes made by the previous queries are already written to disk (checkpointed), semantic security is provided to all the data.

6.2. Security analysis

We will argue that StealthDB satisfies the above leakage profile, by building on the following assumptions:

1) Remote and local attestation provided by SGX are secure according to Section 2.
2) The confidentiality of the intermediate values of the computation and the integrity of the computation from SGX.
3) The confidentiality and integrity provided by the authenticated encryption of AES-GCM.
4) The confidentiality of ElGamal encryption - used during secure channel establishments.

Outline. We will first argue that no information about the master key $K$ is revealed; also that only the permitted clients can make the DBMS execute queries; then use these to argue the semantic security of the data values during the initialization of the database; finally, use all these to argue the leakage during the query execution phase.

Claim 6.1. The confidentiality and integrity of the master key $K$ is ensured throughout the StealthDB execution.

Proof. The database owner forms the root of trust as in Figure 7. The owner is involved in a remote attestation protocol with Auth to check the correctness of the code and the constants loaded into Auth against the publicly available expected measurement of Auth. (The constants loaded into Auth include the expected measurements of PreProcessor and Ops). The master credentials for the database is transferred to a valid Auth. And, the security of SGX remote attestation guarantees the validity of Auth. From this point, the trust is transferred to Auth. Auth generates the master key $K$.

The master $K$ is then transferred to the other enclaves PreProcessor and Ops by Auth through the secure channels established on top of local attestation. The security of local attestation ensures that Auth establishes secure channels with only those PreProcessor and Ops whose measurements match the expected hardcoded ones. Hence, $K$ is transferred only to the correct instances of PreProcessor and Ops. Here, the confidentiality provided by the public key cryptography used in the secure channel establishment (on top of the authenticity from attestation) and the confidentiality and integrity of AES-GCM ensure that no information about $K$ except its length is leaked to an adversary during the transfers.

Now, there are only two more operations which involve $K$. First, when $K$ is used to AES encrypt and decrypt data values, the SGX security guarantees combined with the obliviousness of the AES-NI instructions ensure that no intermediate values about $K$ are leaked. Finally, $K$ is also sealed and stored on the disk for later retrieval. Here, the SGX sealing process provides confidentiality and integrity to $K$. 

\[ \square \]
Claim 6.2. During the query execution phase, a query which reaches the DBMS for execution satisfies the access control policies for the client requesting the query.

Proof. The security of remote attestation also ensures that the database owner transfers the client credentials database only to a valid Auth. When a client proxy initiates a connection with the DBMS, a valid Auth establishes a session with the client only if the client has valid credentials. Next, Auth transfers the session key $s_{esk}$ (shared with the client) only to a valid PreProcessor. This is ensured by the security of local attestation. Now, when the client issues a query, the I/O layer relays it to PreProcessor and PreProcessor parses the query and proceeds only if the query satisfies the access policies of this client. Since there is no other interface for the client to issue a query to the semi-honest DBMS, StealthDB ensures that the semi-honest DBMS only executes a query from a valid client satisfying the access policies provided by the database owner.

Claim 6.3. Semantic security is provided to the data values in a client query before it reaches the DBMS for execution.

Proof. The security of remote attestation ensures that the client provides its credentials and initiates a session only with a valid Auth. The confidentiality and integrity provided by the secure channel ensures that semantic security is provided first to the client credentials, and to the query when the client issues it. Then, PreProcessor parses the query and encrypts the data values. The SGX security ensures that no intermediate values are revealed in this process And AES-GCM ensures the confidentiality of the data values in the output enquery of PreProcessor.

Claim 6.4. The confidentiality and the integrity of a data value is maintained throughout the StealthDB execution unless that specific data value is accessed by a query. In other words, the total leakage of StealthDB is the union of query access pattern leakages.

Proof. All the data values in the database are encrypted using the master key $K$ using AES-GCM encryption. This provides confidentiality and integrity to all the data values. Thus, StealthDB provides semantic security during a steady state when no operations are being performed on the database.

Claim 6.2 lets us focus on the leakage due to running queries from a permitted clients and queries. Let us try to estimate the leakage for a single query. With all the individual data values encrypted (with semantic security) in both the query (from Claim 6.3) and the database, a persistent adversary can only obtain information about the encrypted data values whenever the DBMS makes a data-dependent branching decision. Here, the information learnt by the adversary is exactly the outcome which led to the branching decision. Across multiple executions of Ops, the total leakage is the information that can be obtained from the union of these individual outcomes. The output of the queries also reveal some statistics about the data values, which also form part of our leakage profile.

The above discussion is for starting with an empty database. But, if the data owner would like to upload and work with an existing database, he/she first loads the existing data into the cloud database through Auth only if the remote attestation with Auth is successful. This will thus reduce to the case where we start with an empty database and issue multiple insert queries at the beginning to populate the database, and then proceed with the other steps.

6.3. Security Non-Goals

Integrity of query execution We do not provide integrity guarantees to the clients on the correctness of the query results. This would be an interesting follow-up to StealthDB.

Integrity of index pages We encrypt the index pages and logs only using AES-CTR mode. Since our goal is to protect against passive adversaries, we do not provide cryptographic integrity on the encryption of index pages on disk. Adding an authentication layer of MACs would necessarily involve storing an extra few bytes, and hence an additional I/O management overhead for the DBMS. This is, of course, not an inherent limitation, and this layer can be easily added to the StealthDB I/O layer when encrypting/decrypting pages on disk.

Access pattern to index content Our system leaks the access pattern during query execution in the index pages. In Postgres, the index file stores data as 8 KB pages. When a new value is inserted into the table, only the pages that need to be changed are marked as dirty in the memory and eventually changed on disk. This conforms to the leakage for our system during query execution that we described in Section 6.1.

7. Implementation and Performance

7.1. Implementation details

We implement StealthDB in C and C++ on top of PostgreSQL 9.6 as an extension which loads new SQL objects such as functions, data types, operators and index support methods. To install the extension one can
use command such as `CREATE EXTENSION stealthdb;`. The command loads files `stealthdb.so` (the main library), `enclave_stealthdb.so` (part of the code which is executed in enclaves), `stealthdb.control` (the version control file), `stealthdb.sql` (definitions of new defined functions) to the system. For instance, the function `encintcompare` compares two `encint` values and returns `-1, 0, 1`.

```
CREATE FUNCTION encintcompare(ENCINT, ENCINT) 
RETURNS INT
AS 'libdir/stealthdb'
LANGUAGE C IMMUTABLE STRICT;
```

Figure 12: Example of a new function definition in `stealthdb.sql`

```
PG_FUNCTION_INFO_V1(encintcompare);
Datum
encintcompare(PG_FUNCTION_ARGS)
{
  char *src1, *src2, *pDst;
  int resp, ans = 0;
  char *cl = PG_GETARG_STRING(0);
  char *c2 = PG_GETARG_STRING(1);
  resp = compareInt64(cl, c2, &ans);
  pg_error_handler(resp);
  PG_RETURN_INT32(ans);
}
```

Figure 13: Example of new defined function implementation in `stealthdb.c`

The function `compareInt64` is executed in an enclave. We leverage the native framework of PostgreSQL system to create the extension, so no changes to underlying DBMS were made. We implement a pre-parser in `PreProcessor` to encrypt the data values in queries which helps in avoiding changes to the client drivers (JDBC, ODBC) of the system. Our approach can be extended to other SQL-like database using user-defined functions. But some database systems for instance MySQL do not allow to create an independent stand-alone extension and small parts of the original code that work with data types may need to be changed.

To protect against the side-channel attacks on SGX, we make every operation inside an enclave oblivious by leveraging AES-NI and CMOV instructions. The source code of PostgreSQL 9.6 has more than 700K lines of code while the StealthDB has total about 5000 lines of code and 1500 lines run in enclaves.

7.2. Performance evaluation

To measure StealthDB’s performance, we use a server that has an Intel Xeon E3 3.60 GHz CPU with 8 cores and 16 GB of RAM. For our performance experiments, we measure the throughput and latency of an unmodified PostgreSQL 9.6 and two options of StealthDB using a TPC-C trace. The first option leaves IDs in all TPC-C tables (e.g. `w_id`, `o_w_id`, etc.) unencrypted, and in the second option all columns are encrypted. We choose to evaluate an option where IDs are unencrypted because these are auto-generated at the tables and do not directly contain any sensitive data. Their relationship structure may leak some information, but the performance improvements are significant to consider. The results were received by averaging multiple 1000 second runs with check-pointing turned off. We run experiments varying the amount of clients from 1 to 10 (since only one thread with an enclave for all client connections was used) and choosing the best results. The number of clients can be increased if more threads with enclave are used.

Figure 14 shows the throughput for the TPC-C benchmarking for different scale factors, in this case, StealthDB with unencrypted IDs incurs less than 35% penalty to unmodified PostgreSQL for even large scale. This is sufficient for many real-world transactional systems. Option 2 when all columns are encrypted shows more than a factor of 4 penalty to StealthDB with unencrypted IDs. We use scale factor (W) of 16 to measure TPC-C’s requests latency. The tested database includes nine tables with about 10 million rows total, or almost 2GB of data for PostgreSQL (7GB of encrypted database for StealthDB with unencrypted IDs and more than 10GB for StealthDB with encrypted IDs). Figure 15 shows CDF graphs for each type of TPC-C request.

Table 3 and Figure 16 compare the median and av-
Figure 15: Cumulative distribution functions for TPC-C requests under PostgreSQL and StealthDB. For all graphs the left line (solid) represents PostgreSQL, the second line - StealthDB with unencrypted IDs, the right - StealthDB with encrypted IDs.

<table>
<thead>
<tr>
<th></th>
<th>Median</th>
<th>90th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>PostgreSQL</td>
<td>1.6</td>
<td>5.9</td>
</tr>
<tr>
<td>StealthDB, unencrypted IDs</td>
<td>2.8</td>
<td>7.2</td>
</tr>
<tr>
<td>StealthDB, full encryption</td>
<td>15.2</td>
<td>26.1</td>
</tr>
</tbody>
</table>

TABLE 3: Latency statistics of TPC-C requests, ms

8. Conclusion

Encrypted databases are important when one tries to protect sensitive data from unauthorized users, attackers and administrators in cloud settings. In this work, we build StealthDB – a scalable encrypted database from SGX with full SQL query support. We believe that our encrypted database is the most practical and viable solution for an encrypted database in the near future. Systems based on pure crypto algorithms or FPGAs are very valued, but need a lot more work before they can be considered practical for the cloud usages. However, new schemes and attacks in the field are emerging continuously. We believe it’s important to create an open eco-system where any attack can be considered and tested, and systems can be patched and improved. We are in process of uploading our code to a public repository [Ste] over the next few months, and hope others will be able to contribute to our development efforts.

References


