

A Framework for Building Secure, Scalable, Networked Enclaves

Philipp Winter
Brave Software

Ralph Giles
Brave Software

Alex Davidson
Brave Software

Gonçalo Pestana
Brave Software

Abstract

In 2020, Amazon introduced Nitro enclaves—cloud-based secure enclaves that do not share hardware with untrustworthy code, therefore promising resistance against side channel attacks, which have plagued Intel’s SGX for years. While their security properties are attractive, Nitro enclaves are difficult to write code for and are not meant to be used as a networked service, which greatly limits their potential. In this paper, we built a framework that allows for convenient and flexible use of Nitro enclaves by abstracting away complex aspects like remote attestation and end-to-end encryption between an enclave and a remote client. We demonstrate the practicality of our framework by designing and implementing two production-grade systems that solve real-world problems: remotely verifiable IP address pseudonymization and private telemetry. Our practical experience suggests that our framework enables quick prototyping, is flexible enough to accommodate different use cases, and inherits strong security and performance properties from the underlying Nitro enclaves.

1 Introduction

First introduced in 2015, Intel’s Software Guard Extensions (SGX) technology inspired diverse applications but also increasingly sophisticated attacks: researchers successfully adapted speculative execution attacks [1], injected software faults [2], and exploited side channels introduced by shared caches [3], all with the goal of exfiltrating information that was meant to remain in the enclave. The underlying flaw that most attacks take advantage of is that the untrustworthy operating system and the enclave share a CPU, which provides many options for side channel attacks.

In 2020, several cloud providers began offering “confidential computing” solutions; Google’s is based on AMD’s Secure Encrypted Virtualization (SEV) [4] while Microsoft’s is based on SGX [5]. Both offerings inherit the attack classes that plague their respective architectures. Amazon took a different path by offering a new enclave architecture based on

their Nitro virtual machine isolation technology [6]. Nitro enclaves are separate virtual machines with hardware-enforced CPU, memory, and device isolation, which imposes limits on access by untrustworthy code. While the architecture appears promising, Nitro enclaves remain difficult to use: documentation is sparse, few applications exist, and enclaves can only interact with the parent EC2 instance via a constrained, socket-like interface. This paper presents the design, implementation, and real-world application of a software framework that facilitates the development of networked Nitro enclaves. Key features of our framework include (i) the ability for enclave code to seamlessly and safely access the Internet; (ii) a design for the horizontal scaling of enclaves by synchronizing secret key material; and (iii) a reproducible build system and tooling that allows users to remotely verify an enclave’s authenticity.

During the development of our framework, we had to overcome several challenges. First, Nitro enclaves are meant to be highly constrained environments and therefore lack a dedicated networking interface. We designed a mechanism that allows enclaves to seamlessly send and receive data over the Internet while maintaining an allow list of destinations, for defense in depth in case of enclave compromise. Second, the attestation process for Nitro enclaves was not designed to be performed over the Internet. We developed a way to bind a TLS session to an attestation document to assure clients that they are communicating with an authentic enclave. Third, we had to devise a reproducible and yet easy-to-use build pipeline that allows end users—regardless of their operating system—to compile the enclave application and end up with the exact same image ID as the enclave provider. Fourth, there is no out-of-the-box way for enclaves to scale horizontally while synchronizing their key material. We therefore designed a mechanism that allows enclaves to securely share their key material.

Having overcome the above design challenges, we implemented an easy-to-use Go framework that abstracts away the difficulties and pitfalls of working with networked enclaves. The use of Go allows for rapid prototyping and greatly re-

duces the risk of memory corruption bugs because of Go’s memory safety. We conduct latency measurements to show that our framework can handle high-throughput and real-time applications, and we demonstrate its usefulness and robustness by building two applications on top of it.

Contributions This work makes three core contributions.

- The design and implementation of a freely available Go framework that facilitates the implementation and deployment of enclave applications. The framework consists of a library that an application can use to run as an enclave, and tooling that facilitates deterministic builds and seamless communication with the secure enclave.
- We make it possible via our framework to turn enclaves into networked applications that can easily scale horizontally to respond to increases in load.
- We build two real-world applications on top of our framework, the first of which—a remotely verifiable IP address pseudonymization system—is a contribution in its own right.

Structure Section 2 provides background on secure enclaves in general and AWS Nitro enclaves in particular. Section 3 introduces the design and implementation of our software framework in addition to the build process that guarantees reproducible enclave application builds, followed by Section 4, which presents two production-quality applications that we built on top of our framework. We evaluate our framework in Section 5 and discuss its limitations in Section 6. Finally, Section 7 contrasts our work with past research.

2 Background

This section provides an overview of secure enclaves in general (§ 2.1) and AWS’s implementation in particular (§ 2.2).

2.1 Secure Enclaves

Computers operate on data that is at rest, in transit, and in use. We have well-understood and practical ways to protect data at rest (e.g., full disk encryption) and in transit (e.g., TLS) but only limited solutions for data that is in use. Cryptography provides solutions in the form of fully homomorphic encryption (FHE) and secure multiparty computation (MPC) but for many applications, those building blocks remain too slow and cumbersome. Trusted execution environments—in particular in the form of “secure enclaves”—provide an alternative that is rooted in hardware and code. Unlike FHE

and MPC, enclaves perform at native (or near-native) execution speed because they are general-purpose computing environments that are not limited to the computation of carefully designed functions. Conceptually, enclaves are isolated execution environments that are shielded off from a computer’s main execution environment, which runs the untrustworthy (from the enclave’s point of view) operating system. Enclaves offer various security properties but in the context of this work, we rely on the following three:

Confidentiality An unauthorized entity must not be able to observe the data that an enclave is computing.

Integrity An unauthorized entity must not be able to modify the data that the enclave is computing on, or the code it is running.

Verifiability A separate entity must be able to verify if the enclave is running the code that its operator claims it is running.

Modern CPUs of major hardware vendors implement secure enclaves: Intel has SGX, ARM has TrustZone, and AMD has SEV. A frequent critique of these industry efforts focuses on their proprietary nature. The community has a conceptual understanding of the mechanisms behind enclaves but their exact hardware implementation is not disclosed, which served as motivation towards an open source enclave [7]. In practice, enclaves promise to be useful in situations where a system must process sensitive data while simultaneously be shielded off from the complexity (and subsequent insecurity) of general-purpose computers.

2.2 AWS Nitro Enclaves

In this work, we build on top of AWS’s Nitro enclaves. Nitro enclaves are isolated and constrained virtual machines that run alongside an EC2 instance that is responsible for starting the enclave, and communicating with it. Crucially, an enclave does not share hardware resources with its parent EC2 image; it is guaranteed to have its own CPU and memory which is isolated from the parent EC2 image by the same hypervisor that isolates EC2 instances from each other. As far as computing resources go, Nitro enclaves are essentially an independent computer, with its own operating system, CPU, and memory, but *without* a networking interface or persistent storage. By design, the enclave’s network traffic must go through the parent EC2 instance, constrained to a minimal VSOCK interface [8]. Originally proposed for communication between a hypervisor and its virtual machines, AWS repurposed the VSOCK interface to serve as communication channel between an enclave and its parent EC2 instance. From a developer’s point of view, the VSOCK interface is a point-to-point interface connecting the two. On the address layer, 32-bit context IDs take the role of IP addresses

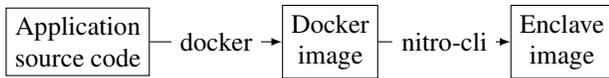


Figure 1: The development workflow for compiling enclave applications. Docker’s command line tool compiles application source code into a Docker image, which is then compiled to an enclave image file using the nitro-cli command line tool.

in VSOCK interfaces. For example, the enclave may have context ID 4 while its parent EC2 instance may have context ID 3. On the transport layer, one can use the same protocols that one can use over the IP-based address family; namely TCP, UDP, etc.

Figure 1 illustrates the development workflow for enclave applications: the workflow begins with the creation of a Docker image that contains the application that will run in the enclave. Using Amazon’s nitro-cli command-line tool, the developer then compiles the container image to an enclave image file (EIF). The compilation process results in a number of *measurement checksums* that uniquely identify the image itself, its kernel, and application. As we will discuss later in the paper, these measurements are key to the remote attestation process. Once the EIF is ready, the developer starts the enclave on an EC2 instance using the nitro-cli command-line tool. The only way for the EC2 instance to exchange data with the enclave is via the VSOCK interface.

3 Framework Design

This section introduces the design of our framework. We start by laying out the involved parties and their respective trust assumptions (§ 3.1), followed by an overview of our system (§ 3.2). We then discuss the two major aspects of our framework: the build system (§ 3.3) and the framework itself (§ 3.4).

3.1 Trust Assumptions

Our setting has three participants that have the following trust assumptions:

1. The *service provider* runs a service for its clients. As part of its operations, the service provider wants to process sensitive client information.
2. The *client* is a user of the service provider. It does not trust the service provider with its sensitive information and demands verifiable guarantees that the service provider will never see the client’s sensitive information in plaintext.
3. The *enclave provider* makes available enclaves to the service provider. Both the client and the service

provider trust that the enclave provider’s enclaves have the advertised security attributes of integrity, confidentiality, and verifiability.

3.2 Design Overview

We begin with a short, informal overview of our system to provide intuition. The subsequent sections are going to elaborate on this high-level picture. The life cycle of an enclave application that uses our framework involves five steps:

1. The service provider implements a new service with the intention of running it in an enclave. Once the implementation is finished, the service provider publishes the source code for its clients to audit, and runs the code in an enclave. After booting, the enclave automatically obtains a CA-signed TLS certificate.
2. Users audit the source code. Once a user has convinced herself that the code is free of bugs, she compiles the code using the framework’s deterministic build system, resulting in an image checksum.
3. The client establishes an end-to-end encrypted network connection with the enclave, facilitated by the EC2 host blindly forwarding encrypted bytes. Right *after* establishing the TLS connection but *before* revealing any sensitive information, the client provides a nonce and asks the enclave for an attestation document.
4. The enclave receives the nonce and asks its hypervisor to generate an attestation document that should contain the client-provided nonce *and* the fingerprint of the enclave’s CA-signed certificate in addition to the usual image measurement checksums. The resulting attestation document is returned to the client.
5. The client performs various checks (see § 3.4.4 for details) and trusts the enclave if all checks pass. The client is then convinced that it’s communicating with the code that the user audited in the previous steps, and is willing to reveal her sensitive information to the enclave.

3.3 The Reproducible Build System

Only a small subset of the users will be skilled enough programmers to audit the enclave’s code for bugs. We expect non-technical users to trust that other users—or perhaps professional code audit companies—have studied the code and pointed out potential bugs. Once a user has convinced herself of the code’s correctness, she compiles the code to arrive at an image ID. Crucially, we need a *deterministic mapping* between the code and its corresponding image ID because the service provider and clients must agree on the image ID that’s running in the enclave.

The popular Docker tool does not offer a deterministic mapping because, among other things, Docker records timestamps in its build process, causing subsequent builds of identical code to result in different image IDs.¹ To obtain reproducible builds, we replace Docker with the kaniko tool [9]. Kaniko’s main purpose is to build container images from a Dockerfile while itself in a container, but we use kaniko because it can do so reproducibly. As long as the client and service provider use the same enclave source code, Go version, and kaniko version, they can build identical images—even when compiling the code on different platforms, like Mac OS and Linux. Equipped with a locally-compiled Docker image ID, the client is now ready to interact with the enclave.

3.4 Framework Components

Having discussed how the client and service provider can independently compile identical Docker image IDs, we now turn to the specific components of our framework. The following sections discuss how the framework seeds its empty entropy pool (§ 3.4.1); how it communicates with the outside world (§ 3.4.2); how it obtains a CA-signed certificate (§ 3.4.3); how we facilitate remote attestation (§ 3.4.4); how enclaves can share their key material to allow for horizontal scaling (§ 3.4.5); how to thwart side-channel attacks (§ 3.4.6), how to ingest secrets (§ 3.4.7), and concludes with a simple example (§ 3.4.8).

3.4.1 Seeding the Entropy Pool

Like many virtual machines, a Nitro enclave is an entropy-starved, sterile environment without any periphery devices that help the kernel seed its entropy pool. To work around that, the Nitro hypervisor can provide randomness for the enclave to seed its entropy pool. Our framework automatically takes advantage of that when it first starts (step ① in Figure 2), so application developers never encounter function calls that block because of insufficient randomness.

3.4.2 Enabling Networking

Recall from Section 2.2 that Nitro enclaves have no networking interface and are only able to communicate with their respective EC2 host. Our framework therefore needs proxying code that runs on the EC2 host and forwards networking packets between clients and the enclave. Figure 2 illustrates the design of the networking architecture.

When the enclave first starts, it fetches a CA-signed certificate from Let’s Encrypt (cf. § 3.4.3) using the ACME protocol [10], which allows for the automated issuance of certificates. To do so, it first connects to Let’s Encrypt’s infrastructure via a SOCKS proxy (step ②). Let’s Encrypt does not

¹In essence, a Docker image is merely a file system. A Docker image is reproducible when separate build processes arrive at the exact same file system, including meta data like timestamps.

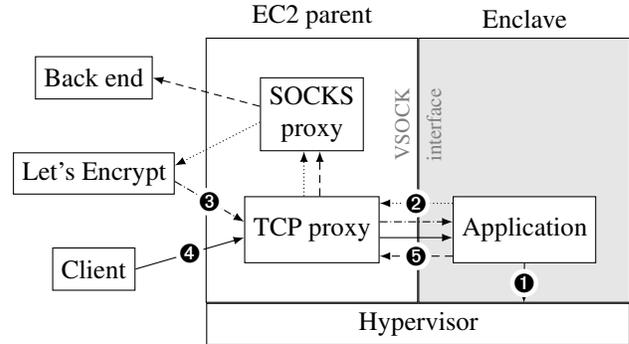


Figure 2: Upon bootstrapping, the application first asks the hypervisor for randomness to seed its entropy pool (①), followed by initiating an ACME session to obtain a Let’s Encrypt-signed certificate (②), after which Let’s Encrypt probes the enclave and issues the certificate (③). Afterwards, clients can establish HTTPS connections with the enclave (④) and the enclave can forward data to its back end (⑤). All of the application’s ingress and egress traffic is routed over a TCP proxy that translates between AF_INET and AF_VSOCK. Egress traffic is reaches the Internet via a SOCKS proxy.

publish its endpoints’ IP addresses, which is why we cannot use point-to-point connections and have to rely on the flexibility of a SOCKS proxy.² Let’s Encrypt then verifies that the enclave can answer queries on behalf of the requested domain name, and issues a certificate to the enclave (step ③).

Clients establish end-to-end encrypted TLS sessions to the enclave with the help of a TCP proxy on the EC2 host that translates between AF_INET (traditional, IP-based sockets) and AF_VSOCK (VSOCK-based sockets). Note that the EC2 host is blindly forwarding encrypted bytes, and cannot see what data the client and the enclave are exchanging (step ④). In an optional, final step, the enclave can send data to a back end (step ⑤).

In case of an enclave compromise, we don’t want the application to be able to leak data to an attacker-controlled endpoint via the SOCKS proxy, which is why we use an allow list on the SOCKS proxy. It is the service provider’s responsibility to maintain the allow list.³ In a minimal application, the allow list consists of two endpoints:

1. The domain name `acme-v02.api.letsencrypt.org` to interact with Let’s Encrypt’s infrastructure.
2. The IP addresses or domain names of whatever back end machine the enclave needs to talk to.

²Our SOCKS proxy is available at <https://github.com/brave-intl/bat-go/tree/nitro-utils/nitro-shim/tools/socksproxy> and our TCP proxy is available at <https://github.com/brave-experiments/viproxy>.

³Note that clients are unable to remotely verify that the service provider correctly configured an allow list because the SOCKS proxy runs on the EC2 host and not in the enclave application.

If an attacker discovered a remote code execution bug in the enclave application and uses the bug to make the enclave establish a connection to any domain name that is not part of the allow list, like `evil.com`, the SOCKS proxy is going to reject the connection.

3.4.3 End-to-end Secure Channel

Having established how the enclave can send and receive network packets, we now turn our attention to secure channels; specifically: how can a client rest assured that it is talking to audited enclave application code, without taking advantage of an existing trust relationship?

The application can register arbitrary HTTP handlers that are accessible to the outside world to process data and provide services. Our framework provides a secure channel based on HTTPS, i.e., we make it possible for a client to establish an HTTPS connection with an enclave in a way that the TLS connection is terminated inside the enclave. Clients can use this channel to access the application without revealing data to third-party observers, or to the service provider, outside of the specific ways the enclave application permits.

Once the enclave has initialized its entropy pool, it obtains an HTTPS certificate that allows clients to establish end-to-end encrypted session with the enclave. Crucially, the HTTPS certificate *lives and dies* inside the enclave and its private key cannot be extracted (or injected) by the service provider because enclaves are sealed at runtime. Our framework allows for both the creation of a self-signed or a CA-signed certificate. If a self-signed certificate is desired, the framework creates and signs a certificate for a given FQDN. To get a CA-signed certificate, the framework uses Let's Encrypt's ACME protocol because it allows for the generation of a certificate with no human interaction. In that case, the enclave initiates an HTTP-01 challenge⁴ connection with Let's Encrypt's infrastructure via our SOCKS proxy (cf. Figure 2), and subsequently expects an incoming connection from Let's Encrypt to port 80, which the EC2 host forwards to the enclave. Note that the EC2 host can also obtain a CA-signed certificate for the same FQDN because the enclave and the EC2 host share an IP address. This is however of little use to the EC2 host as we will show in the next section.

3.4.4 Remote Attestation

By default, Nitro enclaves only allow for local attestation. We now discuss how we allow clients to remotely retrieve and verify an enclave's attestation. After the client establishes an HTTPS connection with the enclave, it needs to know that (i) the TLS connection it just established is terminated inside the enclave (instead of by the EC2 host) and

⁴ACME supports multiple challenge types, with HTTP-01 being the most common one [11]. In HTTP-01, the ACME infrastructure provides the client with a token, which must subsequently be available at a pre-defined path on the client's Web server.

(ii) the enclave is running the code that the user audited in the previous step. To that end, the client requests the enclave's *attestation document*—a hypervisor-signed document that attests to the container image ID that the enclave is running. Enclaves communicate with the hypervisor via the `ioctl` system call, which makes use of `/dev/nsm`, a device that is only available inside a Nitro enclaves. To request an attestation document, the client provides a *nonce*—a 160-bit random value—whose purpose is to prevent the service provider from replaying outdated attestation documents. Phrased differently, the client provides a nonce to convince itself that it's talking to a live enclave. In practice, clients make the following HTTP request to obtain an enclave's attestation document:

```
GET /attestation?nonce=8083...23b7 HTTP/1.1
```

The enclave receives the request through the TCP proxy, asks the hypervisor to include the nonce *and* the fingerprint of the enclave's X.509 certificate in the attestation document, and sends the resulting attestation document to the client. By asking the hypervisor to include the certificate fingerprint in the attestation document, we effectively bind a TLS session to an enclave, which is key to assuring the client that it is in fact talking to an enclave. Upon receiving the attestation document, the client then verifies the following in order:

1. The attestation document is signed by the AWS PKI whose public key is known to all parties.
2. The challenge nonce is part of the attestation document.
3. The fingerprint of the enclave's X.509 certificate from the TLS session is part of the attestation document.
4. The enclave's image ID is identical to the image ID that the client compiled locally.

Only if all four conditions hold is the client convinced that it is talking to an enclave running the code that the client audited, and that the TLS connection is terminated by the enclave. Note that the EC2 host is able to intercept HTTPS connections with its own CA-signed certificate but clients will only trust the EC2 host if (and only if) it can present an attestation document that is valid for the enclave image, which it can't because it is unable to spoof the AWS PKI signature that authenticates the attestation document. The only way for the EC2 host to obtain such an attestation document is to spawn an enclave that runs the exact code that the client is expecting—and it already is doing exactly that. Now that the client has established a trust relationship with the enclave, it is ready to reveal sensitive information to the enclave.

While attestation documents can be generated quickly and in rapid succession (cf. § 5.3), they do require an extra round trip between the client and the enclave before the client is willing to reveal sensitive information: the client first provides a nonce to the enclave, the enclave responds with an

attestation document, and only after the client verified the document is it willing to reveal its sensitive information. To eliminate unnecessary future round trips, clients should use TLS session resumption once they have verified the enclave’s attestation document. The service provider is unable to see the secret key material that protects the TLS session between client and enclave, so it is safe to re-use a once-established TLS session and forgo the unnecessary round trip.

Manual Client-side Verification In general, we envision that remote attestation is handled transparently by client software, without involving the user. In some cases however, users may wish to manually verify an enclave. Even for developers, remote attestation is a complex process that is difficult to understand and work with. To make matters more complex, in our setting end users are expected to conduct remote attestation. We therefore made a careful effort to abstract away technical details. A user wishing to remotely verify an enclave essentially asks herself “does the enclave that’s exposed at a given URL run the source code that I just audited?” We built a tool set that reduces the process to the running of a Makefile,⁵ i.e.:

```
$ make verify CODE="/path/to/enclave/code/" \
    ENCLAVE="https://example.com/attest"
```

The first argument, CODE, points to the directory containing the source code that the enclave is supposedly running, and the user audited. The second variable, ENCLAVE, points to the URL endpoint of the enclave that the user’s client is connecting to. When the user runs this command, the Makefile deterministically compiles the given source code to obtain its image ID, asks the enclave for an attestation document, verifies the document, and ensures that the attestation document is for the image ID that was compiled in the first step. If all checks pass, the tool informs the user accordingly.

3.4.5 Sharing Key Material

Recall that enclaves are essentially sealed at runtime, preventing anyone (including both Amazon and the service provider) from extracting key material that was generated inside the enclave. While this is a desirable property, it complicates horizontal scaling. If a single enclave is unable to handle the service provider’s traffic load, one must scale horizontally by starting new enclaves. In some applications, it is unacceptable for each enclave to use distinct key material. Instead, enclaves must synchronize their key material, so they appear to the outside world like a single machine.

While it is possible to build key synchronization using tools like the AWS key management service (KMS),⁶ we re-

⁵The source code is available at: <https://github.com/brave-experiments/verify-enclave>.

⁶One could encrypt the keys using a KMS policy that dictates that only enclaves are allowed to decrypt it, and store the encrypted key in a location that all enclaves can access, e.g., an S3 bucket.

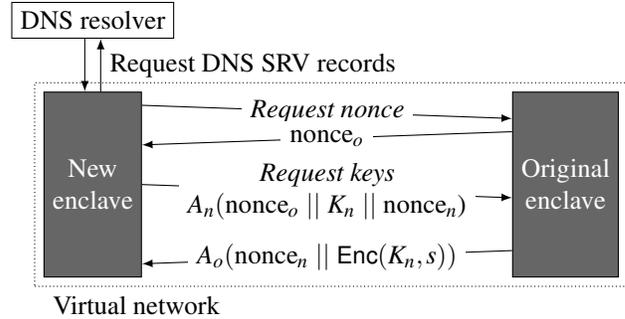


Figure 3: When a new enclave bootstraps, it discovers existing enclaves by obtaining the DNS SRV record for its own, hard-coded FQDN. The enclave then initiates key synchronization by first requesting a nonce. Then, the new enclave requests the origin enclave’s key material by submitting its own attestation document, followed by receiving the origin enclave’s attestation document, which contains encrypted key material.

frain from using KMS because there is no straightforward way for users to verify that the service provider is using KMS as promised. We therefore devise a new protocol that enables key synchronization without having to rely on external services.

We solve this problem in two steps: *discovery* and *synchronization*. First, enclaves must be able to discover each other, i.e., learn each other’s IP addresses. Then, enclaves can establish connections to each other and initiate key synchronization. Our protocol dictates that when a new enclave bootstraps, it first tries to discover already-existing enclaves. If there are none, the enclave knows that it is the “origin” enclave; it generates new key material and is prepared to share it with future enclaves. If however it discovers other enclaves, the new enclave establishes a connection to another randomly-chosen enclave and initiates key synchronization. Crucially, key material is only shared after *mutual attestation*, i.e., the original and subsequent enclaves verify each other, and only exchange key material if remote attestation succeeds. Key synchronization happens in three steps, as illustrated in Figure 3.

1. Once a new enclave is spun up, it queries the DNS SRV record of the FQDN that is hard-coded in the enclave, e.g., example.com. The DNS resolver will return the record, containing a list of enclaves that are already running and initialized. The new enclave picks a random enclave from the list and initiates key synchronization.
2. The new enclave asks the existing enclave for a random nonce, nonce_o. Each enclave caches nonce_o for one minute.
3. The new enclave now requests the key material from the existing enclave. As part of the request, it provides

its attestation document that contains nonce_o (to prove freshness to the existing enclave); nonce_n (the existing enclave is expected to add this nonce to its attestation document); and K_n (a public key to which the key material should be encrypted). Upon receipt of the new enclave’s attestation document, the existing enclave verifies the attestation document’s signature and ensures that the new enclave is running the same code, i.e., the measured checksum value that uniquely identifies the enclave image is identical. Once the existing enclave is convinced that it is dealing with a genuine new enclave, it creates an attestation document by including nonce_n (to prove freshness to the new enclave) and $\text{Enc}(K_n, s)$ —the key material s is encrypted using the public key that the new enclave provided in the request. Finally, the new enclave verifies the attestation document, decrypts the key material, and uses it to finish bootstrapping.

Needless to say, the security of key synchronization is paramount. As an optional first layer of defense, synchronization should be configured to use a private network segment limited to inter-enclave communication, such as the internal network that is part of a Kubernetes cluster that is managing scaling for the application. While optional, Kubernetes is an attractive component in our setting considering that enclaves are essentially compiled Docker images. The second and main layer of defense is the fact that an enclave has to provide a valid attestation document before the origin enclave reveals its key material. As long as the origin enclave knows that an identical and authentic copy of itself is asking for key material, it will readily provide it.

3.4.6 Side-channel Attacks

The enclave’s parent EC2 host cannot see *what* clients send to the enclave but it can see *how much* clients send and *how long* it takes the enclave to process data. The EC2 host can exploit these side channels to learn more about the client’s confidential information and computation. While such side channels must be avoided, our framework is not the place to do so. Instead, it is the application developer’s responsibility to identify and address this class of attacks. Section 4 introduces two applications and discusses side channel attacks in their respective setting.

Similarly out of scope are programming bugs in the enclave application. Memory corruption bugs may be more difficult to exploit in an enclave application⁷ but Lee et al. adapted a return-oriented programming attack against SGX to show that such attacks are practical [12].

⁷The untrustworthy operating system (that may be under the attacker’s control) is prohibited by hardware to read the enclave application’s memory or registers in clear text, which forces the attacker to operate blindly.

3.4.7 Ingesting Secrets

A key design requirement of our framework is that users can audit and verify the code that is running inside an enclave, which means that the service provider is unable to hide any software configuration from the user. Service providers can work around this shortcoming by implementing HTTP handlers that take as input arbitrary data, and use it to update the enclave’s state. Consider a system that takes as input client IP addresses, anonymizes them, and forwards the anonymized addresses to a back end (cf. § 4.1). The service provider now wants to compare submitted IP addresses to a confidential deny list. However, if the deny list is hard-coded in the freely available enclave application, it is readily visible to anyone. The service provider can solve this problem by adding to the enclave application a new HTTP handler that takes as input the confidential data it seeks to protect from the users’ eyes. Once the enclave is running, the service provider loads the confidential data at runtime, by calling the end point. To prevent users from submitting bogus data, the endpoint could hard-code the service provider’s public key and only accept data that carries a valid signature of the service provider’s private key.

This technique for ingesting secrets into an enclave’s runtime is flexible—so flexible, in fact, that the service provider could abuse it to ingest code at runtime, which would nullify the enclave’s verifiability requirement. Vigilant users would never trust an enclave whose code can change at runtime. We therefore argue that an HTTP handler for the purpose of ingesting secrets must be constrained to a point that only data of a well-defined type can be ingested.

3.4.8 An Example

Figure 4 illustrates an example of a simple “hello world” application. The code initializes a new enclave struct (line 16), followed by adding a handler that processes requests for GET /hello-world (line 24). Finally, the application starts the enclave using a function call that does not return (line 27).

The configuration object (line 17) consists of four fields. The first field determines the enclave’s FQDN, which is required when obtaining an X.509 certificate. The port is listened on by the enclave via the VSOCK interface. It is the EC2 host’s responsibility to forward incoming traffic to this port. The third field determines if the enclave should use Let’s Encrypt’s ACME protocol to obtain a CA-signed certificate at runtime. The last flag instructs the framework to print debug information. Note that this is only useful when the enclave is invoked in debug mode, which disables remote attestation.

```

1 package main
2
3 import (
4     "fmt"
5     "log"
6     "net/http"
7
8     nitro "REDACTED"
9 )
10
11 func handler(w http.ResponseWriter, r *http.Request) {
12     fmt.Fprintln(w, "hello world")
13 }
14
15 func main() {
16     enclave := nitro.NewEnclave(
17         &nitro.Config{
18             FQDN:    "example.com",
19             Port:    8080,
20             UseACME: true,
21             Debug:   false,
22         },
23     )
24     enclave.AddRoute(http.MethodGet,
25         "/hello-world",
26         handler)
27     if err := enclave.Start(); err != nil {
28         log.Fatalf("Terminated: %v", err)
29     }
30 }

```

Figure 4: An example of a simple enclave application which registers an HTTP GET handler for the path `/hello-world` (line 24) and, when accessed, responds with the string “hello world” in the response body (line 12).

4 Applications

We demonstrate the practicality of our framework by building two applications on top of it. First, in a novel system that pseudonymizes client IP addresses for anti-fraud (§ 4.1); second, to implement the shuffler component that is part of Bittau et al.’s PROCHLO paper [13] (§ 4.2). In both cases, we had to overcome minor obstacles but found that our framework greatly facilitated the deployment of enclave code.

While our framework is written in Go, developers can build enclave applications in languages other than Go, by taking advantage of a foreign function interface (FFI) that allows Go code to invoke functions in other languages. We successfully moved a Rust code base into a Nitro enclave by implementing a lightweight, Go-based wrapper (in less than 200 lines of code, excluding our framework’s code) that interacts with the Rust code via an FFI.⁸ Importantly, the compilation process is still reproducible as long as the Rust code’s dependencies are pinned via Rust’s Cargo tool chain.

⁸The code is available at: <https://github.com/brave-experiments/star-randsrv>.

4.1 IP Address Pseudonymizer

Our first application is the system that originally motivated us to build the enclave framework. Consider a service provider that offers various services to its users. The service provider seeks to know as little as possible about its users, which means that it doesn’t capture and store any of its users’ IP addresses. IP addresses are however an important signal in the service provider’s fight against a subset of its users that commit fraud. This constitutes a conundrum: Should the service provider collect all its users’ IP addresses to strengthen its anti-fraud efforts? Or continue to discard the addresses, and tolerate the fraud?

This section presents an application that strikes a balance between these two extremes; an IP address pseudonymization system that can verifiably pseudonymize IP addresses. The service provider can then run its anti-fraud logic over pseudonymized IP addresses rather than real ones. While some information is lost in the process, we argue that what’s most important—the relationship between IP addresses—can be preserved thanks to our use of the Crypto-PAn scheme that Xu et al. presented in their 2001 paper [14]. Crypto-PAn encrypts both IPv4 and IPv6 addresses by implementing a 1:1 mapping f that is keyed by k from an IP address to its pseudonymized equivalent while *preserving the address’s prefix*, i.e., two IP addresses that share an n -bit prefix also share an n -bit prefix after pseudonymization as illustrated by the following example:

$$f(k, \text{"10.0.0.1"}) = \text{"242.32.192.193"} \quad (1)$$

$$f(k, \text{"10.0.0.123"}) = \text{"242.32.192.154"} \quad (2)$$

Figure 5 illustrates the system design. Clients periodically communicate with a service behind a reverse TLS proxy whose job—among other things—is to hide client IP addresses from the service. The proxy is configured to mirror incoming client requests to the enclave, but with client IP addresses intact, in the form of a custom HTTP header like `X-Client-Addr: 1.2.3.4`. The service is interactive, and responds to the client, but the enclave is passive, and simply consumes the requests. Recall that the (untrusted) EC2 host that hosts the enclave is unable to see the client’s IP address because the TLS proxy establishes a TLS session that’s terminated *inside* the enclave. The pseudonymizer takes as input client requests, extracts the IP address that the proxy inserted from the HTTP header, pseudonymizes them, and forwards pseudonymized addresses in batches to the configured back end. Components in dark gray are under the service provider’s control.

One problem however remains: how do clients know that the TLS proxy is in fact configured to discard client IP addresses before forwarding requests to the service? Unfortunately, cloud providers don’t offer satisfying solutions for this problem but some cloud providers allow for the creation

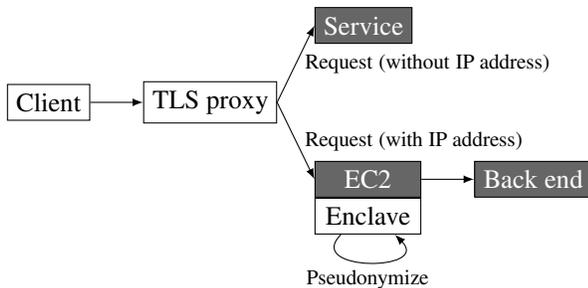


Figure 5: Clients communicate with a service that’s available behind a third-party TCP proxy whose purpose is (among other things) to drop client IP addresses, so the service never sees them. The proxy is configured to mirror incoming client requests *with* IP address to the enclave, where addresses are pseudonymized and finally forwarded to a back end for analysis.

of roles whose permissions are configurable. The service provider can create a read-only role in the TLS proxy’s configuration interface and publish the credentials for this role. Doubtful users can then log in to this role and verify that the TLS proxy is configured to discard IP addresses when forwarding requests to the service. As long as the TLS proxy operator is a neutral third party with no incentive to lie, which is typically the case, both the client and service provider can trust it. We acknowledge that this is not an elegant solution but a mere hack, to work around the shortcoming of Nitro enclaves not having a networking interface. If enclaves had a networking interface that could not be monitored by the untrusted EC2 host, clients could directly talk to the enclave, obviating the need for a TLS proxy that hides client IP addresses from the EC2 host.

Side channels The untrusted EC2 host never sees client IP address in plaintext but it can exploit timing and volume side channels to infer information about the encrypted requests that the TLS proxy forwards to the enclave. We close this side channel by adding code to the enclave application which queues pseudonymized IP addresses until two conditions are true: (i) we have at least n pseudonymized addresses, and (ii) at least t minutes have passed.

Key rotation A single pseudonymous IP address without context cannot be reversed and reveals nothing about its corresponding plaintext IP address but that changes if the service provider expects the client to repeatedly report its IP address to the enclave. For example, a sequence of pseudonymized IP addresses can reveal either that (i) the client has not changed its IP address, or (ii) the client changed its IP address but is likely to use the same ISP (e.g., if the /24 prefix remains the same), or (iii) the client changed IP addresses *and* ISPs (e.g., if the prefixes of the pseudony-

mous IP addresses share less than, say, eight bits). While this is useful information for anti-fraud operations,⁹ it also reveals information about a given client’s location, and service providers may want to err on the side of privacy instead. We therefore added a mechanism for periodic key rotation, so a given client’s pseudonymized IP addresses are only meaningful within a given rotation period. According to the results of Padmanabhan et al., we believe that a key rotation period of three weeks strikes a useful balance between privacy for the client and usefulness for the service provider [15, § 3.2]: several ISPs re-assign many of their users’ IP addresses in less than—or up to—two weeks.

In the final step, the enclave submits the client’s pseudonymized IP address and a hash of the key to the service provider’s back end, where anti-fraud logic is implemented. The implementation details of both the back end and its anti-fraud logic are beyond the scope of this paper.

Implementation Our pseudonymization service counts approximately 1,000 lines of code, including comments and tests. It is important to keep the source code small for both security and transparency: a large code base is more likely to have security-critical bugs and is also more difficult for users to audit.

Alternative pseudonymization In addition to Crypto-PAn, we implemented a second pseudonymization method that is based on HMAC-SHA-256. Like Crypto-PAn, the HMAC is keyed by a 160-bit secret that the enclave generates when first bootstrapping. Unlike Crypto-PAn however, the HMAC-based method does not preserve the prefixes of IP addresses: two IP addresses that differ in only a single bit will result in entirely different hashes. On the privacy/utility spectrum, the HMAC-based method therefore leans more toward privacy.

4.2 k -anonymity Enforcer

Bittau et al.’s SOSP’17 paper [13] proposes a private analytics system that helps service providers gain insight into their clients’ usage patterns. Their system—called PROCHLO—consists of three components: (i) software running on the client, which can (but doesn’t have to) add local differential privacy to client measurements. Measurements consist of product-relevant answers to questions like “has the user interacted with her browser in the last 24 hours?” These measurements are then sent to (ii) a shuffler, which enforces a configurable k -anonymity threshold on incoming measurements. The shuffler then sends measurements meeting that threshold to (iii) an analytics system that the service provider

⁹For example, a service provider would deem a client that constantly changed ISPs suspicious; it is likely to use proxies to connect to the service provider’s infrastructure.

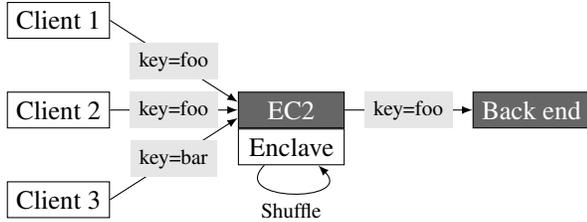


Figure 6: A conceptual overview of our shuffler implementation. Clients send measurements to the shuffler, which enforces a k -anonymity threshold—in this case for $k=2$. Only one of the two measurement types passes the threshold and is forwarded to the back end.

uses to explore users’ anonymized data. The PROCHLO paper envisions the shuffler running in a secure enclave; otherwise users would have no reason to trust that the shuffler is in fact enforcing k -anonymity thresholds. To that end, the authors designed the shuffler to run inside an Intel SGX enclave, which was challenging considering the memory constraints that SGX imposes. For more details about the shuffler, refer to the original PROCHLO paper [13, § 3.3].

As part of an unrelated research project, we were experimenting with private telemetry, and we therefore used our framework to re-implement the shuffler in approximately 1,000 lines of code.¹⁰ Our implementation is a near-complete clone of the shuffler as it was proposed in the PROCHLO paper but for simplicity, we did not implement nested encryption [13, § 3]. Figure 6 shows that our shuffler takes as input confidential client measurements and enforces a configurable k -anonymity threshold on those measurements. Every t seconds, the shuffler discards messages that don’t meet the threshold and forwards the remaining messages to its back end.¹¹ Before clients agree to sending their sensitive measurements to the shuffler, they audit its source code and perform remote attestation, to convince themselves that their measurements are processed by an authentic enclave.

Compared to Bittau et al.’s original, SGX-based design, our Nitro-based implementation has two key advantages: (i) Nitro’s underlying hardware isolation makes our implementation more robust to hardware side channel attacks and (ii) Nitro doesn’t suffer from the same resource constraints as SGX, which renders our implementation easier to use and less error-prone.

Side channels The untrustworthy parent EC2 host can take advantage of the same side channel as with the IP address pseudonymizer. In particular, the EC2 host can observe when

¹⁰The code is available at: <https://github.com/brave-experiments/p3a-shuffler>.

¹¹The variable t depends on the rate of incoming measurements. Reasonable values can range from hours (if the shuffler constantly sees a high rate of incoming measurements) to days.

and how many requests clients make. Like with the IP address pseudonymizer, the PROCHLO paper closes this side channel by aggregating requests, preventing the EC2 host from linking incoming to outgoing requests.

5 Evaluation

We evaluate our enclave framework with respect to security (§ 5.1), financial cost (§ 5.2), and performance. As for performance, we study the rate at which one can generate attestation documents (§ 5.3) and measure end-to-end request latency and throughput (§ 5.4).

5.1 Security Considerations

There are three key components to the overall security of enclave applications; (i) Amazon’s Nitro enclave system itself, (ii) our framework, and (iii) the application that runs on top of our framework.

The very foundation of our framework’s security lies in the soundness of the design of Nitro enclaves. While Amazon published the conceptual design, the concrete hardware and software implementation remains confidential. The decision to allocate physically separate resources to enclaves appears promising but only time will tell if Nitro enclaves can resist the types of attacks that have been plaguing SGX. If we assume that Nitro enclaves are acceptably secure, the next critical layer is our software framework.

A significant security aspect of our framework is its size; it is well understood that complexity is the enemy of security. Our framework counts less than 700 lines of code and has four direct dependencies that are not maintained by either us or the Go project.¹² Four is worse than zero, but is still manageable and reasonably easy to audit in its entirety. We believe that our choice of using Go and the deliberately small trusted computing base greatly reduces—but does not eliminate!—the attack surface.

The highest layer in the software stack is the enclave application itself. The biggest security threat are side channel attacks and programming bugs—both unintentional and intentional. It is the application developer’s responsibility to prevent side channel attacks and write bug-free code. As we pointed out in Section 6, programming bugs can be intentional, i.e., the service provider may deliberately introduce bugs that leak sensitive information. From the user’s point of view, eternal vigilance is therefore the price of security.

¹²The dependencies are chi [16] (provides an HTTP request router), nsm [17] (provides an interface to interact with the Nitro hypervisor), vsoc [8] (provides an API for the VSOCK address family), and tenus [18] (provides an API to configure Linux’s networking devices).

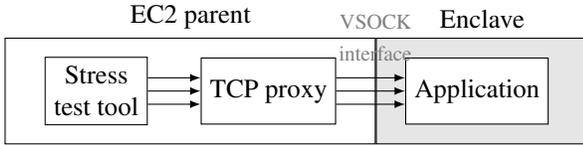


Figure 7: Our stress test tool tests the performance of our critical path, consisting of the TCP proxy, the VSOCK interface, and Go’s HTTP stack in the enclave application.

5.2 Financial Cost

Nitro enclaves do not incur any extra cost in addition to what the underlying EC2 host costs—they can be considered a “free” extension to EC2. Nitro enclaves are however only available for select types of EC2 instances because they require their own CPU and a minimum amount of memory, and those instance types are pricier than the lowest tier that AWS offers.

We are currently working on deploying the IP address pseudonymization prototype that we introduced in Section 4.1. We estimate that our enclave is going to have to handle an average of 5,000 requests per minute, coming from more than ten million clients. Our test deployment uses a single c5.xlarge EC2 host in the U.S. East region which costs \$0.17 per hour to operate, amounting to approximately \$125 per month.

5.3 Attestation Documents

The fetching of attestation documents is a critical part of our framework’s overall performance. We wrote a stress test tool that requests as many attestation documents as it can over sixty seconds. The tool is essentially a minimal enclave application that requests attestation documents in a loop. For each attestation document, we asked the hypervisor to include an incrementing nonce, to avoid any speedups by caching. We were able to obtain approximately 900 documents per second, with each request taking a median of one millisecond ($s = 0.3$ ms) to fetch the attestation document.¹³

5.4 Application Latency and Throughput

Next, we set out to measure the networking latency of the critical path, as illustrated in Figure 7. In particular, we test the latency of our TCP proxy, the VSOCK interface between EC2 and enclave, and a minimal enclave application. We measure latency in three separate setups, designed to help us understand how much latency each component in our data flow adds:

¹³We performed our measurements on a c5.xlarge EC2 host which comes with four CPUs and eight GiB of memory.

Setup	Reqs/sec	Mean lat. (ms)	Max lat. (ms)
Full	7,500	12.7	56.0
No proxy	14,100	6.5	52.0
Direct	27,900	3.2	50.0

Figure 8: Using 100 concurrent requests and 100,000 requests in total.

Full: This represents the full data flow as it would occur in production, i.e. client → TCP proxy → VSOCK interface → enclave application.

No proxy: This setup does not contain the TCP proxy, i.e., the client talks to the VSOCK interface directly, i.e. client → VSOCK interface → enclave application.

Direct: This setup does not contain the TCP proxy and the VSOCK interface. Instead, the client directly talks to an application instance that is running *outside* the enclave, i.e., client → application.

As part of our measurement setup, We first deploy the code from Figure 4—a minimalistic application that responds with the string “hello world” upon receiving requests for the path /hello-world. It’s important to use a minimalistic application because we’re only interested in the latency that is caused by the components *before* a request reaches the enclave application.

To simulate clients, we use the HTTP load test tool Baton [19]. We run Baton on the parent EC2 host and instruct it to send as many requests to the TCP proxy as possible within 30 seconds, using 50 concurrent threads. We had to patch Baton’s source code to add VSOCK support (to be able to send requests directly to the enclave via the VSOCK interface) and to log latency percentiles. Note that our measurements constitute a *lower bound* of the latency that is achievable. Real-world applications will exhibit higher latency because clients send their requests over the Internet (which adds considerable networking latency) and the enclave application is likely to be more complex (which adds computational latency).

Figure 8 illustrates the results for our three test setups. The full pipeline is able to sustain 7,500 requests per second, with a mean latency of 12.7 milliseconds. Removing the proxy nearly doubles the requests to 14,100 per second and lowers the mean latency to 6.5. Finally, a direct connection to the application—without proxy and VSOCK interface—once again nearly doubles the number of requests, reaching 27,900 per second, with a mean latency of only 3.2 milliseconds. Figure 5.4 shows the empirical CDF of the same latency measurements for our three test setups.

Next, we measure the throughput that we can achieve over the VSOCK interface. To that end, we use a VSOCK-enabled fork of the iperf3 performance measurement tool [20]. iperf3

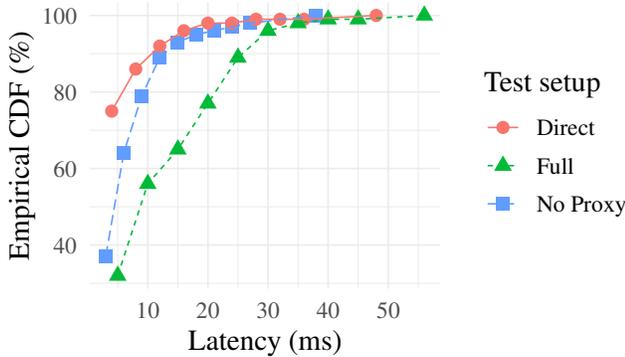


Figure 9: The empirical CDF of the latency distributions of our three test setups.

measures the throughput of a networking link using a client/server model. In our experiment, we start an iperf3 server instance inside the enclave and the corresponding client instance on the parent EC2 host.¹⁴ The client then talks to the server via the VSOCK interface and determines the maximum possible throughput. In this setup, iperf3 measured a throughput of 4.09 GBit/s. For comparison, when running both the iperf3 client *and* server on the EC2 host—which effectively measures the throughput of the EC2 host’s loop-back interface—we achieve 55.5 GBit/s of throughput.

To develop intuition on the perceived network performance of Nitro enclaves, we built an enclave application that acts as a SOCKS proxy. We then configured a browser to use this enclave-enabled SOCKS proxy and browsed HD videos on YouTube. We found that the experience was seamless: videos loaded quickly, played smoothly, and there was no perceivable latency impact when browsing the Web. We believe that the high throughput and low latency, coupled with this anecdotal user experience report suggests that our framework is suitable for demanding and latency-sensitive networking applications.

6 Limitations

We conclude the discussion of our framework by summarizing its limitations.

An obvious limitation is the reliance of our framework on Amazon, which acts as the root of trust. We mentioned in Section 3.1 that all parties must trust Amazon. Note that this is not a new limitation of secure enclaves—SGX-based applications must trust Intel while TrustZone-based applications must trust ARM. Despite the lack of alternatives, placing one’s trust in a single corporation’s proprietary technology is problematic.

¹⁴The command that we ran on the server was “iperf3 --vsock -s” and on the client “iperf3 --vsock -c 4.”

Our system fundamentally relies on at least some users auditing the service provider’s application that runs in a secure enclave. Needless to say, not all users have the skills to audit the service provider’s application and convince themselves that the code is sound. In fact, even among the subset of users that are programmers, only a fraction may feel comfortable auditing source code for vulnerabilities. So what are the non-programmers to do? We envision users to congregate in forums where matters related to the service providers are discussed. A tech-savvy subset of the users is going to organize code reviews and make public their findings. Non-technical users may then trust other users that audited the source code, but this is no different from most other software: nobody audits all the software that they use, ranging from the kernel to the myriad of user space applications, even when source code is available.

The Underhanded C Coding Contest [21] was about implementing benign-looking code that was secretly malicious. The contest attracted numerous impressive submissions which showed that it can be surprisingly difficult to find bugs *even if one knows* that there is a bug in a given piece of code. Analogously, the service provider could try to hide subtle, yet critical bugs in the code to exfiltrate information from the enclave. On top of that, if the service provider ever gets caught, it may have plausible deniability and pretend that the exfiltration bug was an honest programming error. While we are unable to solve this class of attacks, we can mitigate it by keeping the trusted computing base in the enclave as small as possible.

7 Related Work

Arnautov et al. present in their OSDI’16 paper a mechanism that allows Docker containers to run in an SGX enclave [22]—conceptually similar to Nitro enclaves, which are effectively compiled Docker images. In their 2022 arXiv report, King and Wang [23] propose HTTPPA—an SGX-based extension to HTTP that makes a Web server attestable to clients. Our framework also allows for attestable Web services, but without modifications to HTTP.

Applications of Enclaves Researchers have proposed numerous and diverse enclave-enabled systems, ranging from DeFi oracles [24], to health apps for COVID-19 [25], to networking middleboxes [26]. Despite avid interest in academia, large-scale, real-world deployments of enclaves are sparse. In 2017, the Signal secure messenger published a blog post on private contact discovery [27], which makes it possible for Alice to discover which of the contacts in her address book use Signal without revealing her contact list. The Signal team accomplished this by relying on an SGX enclave that runs the contact discovery code. Two years later, in 2019, the Signal team built its “secure value recovery” feature on SGX as

well [28].

Frameworks for Enclave Development To facilitate working with enclaves, several frameworks have emerged that abstract away complicated and error-prone low-level aspects of enclaves. Examples are Asylo [29] and Open Enclave [30]—both libraries are implemented in C/C++ and are hardware agnostic, meaning that the “enclave backend” can be switched from, say, TrustZone to SGX. While frameworks render enclave development more convenient, memory unsafe languages like C and C++ make it dangerously easy to introduce memory corruption bugs that jeopardize the security of the enclave [12]. Cognizant of this issue, Wang et al. implemented a performant Rust layer on top of Intel’s C++-based SGX SDK, making it possible to develop memory-safe applications in SGX [31].

Our framework is built in the memory-safe Go programming language, which eliminates an entire class of bugs that could jeopardize the security of enclave applications, and unlike Asylo and Open Enclave, our framework only supports Nitro enclaves because the security guarantees of a framework are only as strong as the underlying enclave hardware, and in the case of Intel, ARM, and AMD, side channel attacks are a severe concern.

Attacks Against Enclaves Enclaves based on Intel’s SGX technology share a CPU with untrusted code, which raises the flood gates for side channel attacks. Consequently, attacks have taken advantage of speculative execution [1, 32], branch “shadowing” [33], the interface between SGX and non-SGX code [34], software faults [2], shared caches [3], and memory management [35]. Despite the considerable number of practical attacks, there is opportunity to strengthen SGX against side channel attacks. Oleksenko et al. introduce in their ATC’18 paper a system that protects unmodified SGX applications from side channel attacks by executing the enclave code on a CPU separate from the untrusted code. Note that this is the default for Nitro enclaves.

For a comprehensive overview of attacks against SGX, refer to Fei et al.’s survey [36] and Nilsson et al.’s arXiv report [37].

Among all currently-available commodity enclaves, Intel’s SGX has received the most attention from academia but ARM’s TrustZone and AMD’s SEV have not been spared and share SGX’s conceptual security flaws. In a CCS’19 paper, Ryan demonstrates an attack that exfiltrates ECDSA private keys from Qualcomm’s implementation of a hardware-backed keystore which is based on TrustZone [38]. Similarly, Li et al. showed in a USENIX Security’21 paper how an attacker can exfiltrate private keys from AMD SEV-protected memory regions. In a CCS’21 paper, Li et al. showed how an attacker-controlled VM can read encrypted page tables, and how an attacker can create an oracle for encryption and decryption.

While Nitro enclaves are still young and have received nowhere near the same scrutiny as SGX and friends, we believe that their dedicated hardware resources provides stronger protection from side channel attacks than enclaves that are based on shared CPU resources.

Resources

The following list has URLs to all code repositories that are mentioned throughout work.

- Enclave framework:
<https://github.com/brave-experiments/nitriding>
- Remote attestation tool set:
<https://github.com/brave-experiments/verify-enclave>
- TCP proxy:
<https://github.com/brave-experiments/viproxy>
- SOCKS proxy:
<https://github.com/brave-intl/bat-go/tree/nitro-utils/nitro-shim/tools/socksproxy>
- IP address pseudonymization system:
<https://github.com/brave-experiments/ia2>
- PROCHLO shuffler:
<https://github.com/brave-experiments/p3a-shuffler>
- Go application using Rust FFI:
<https://github.com/brave-experiments/star-randsrv>

References

- [1] Jo Van Bulck et al. “Foreshadow: Extracting the Keys to the Intel SGX Kingdom with Transient Out-of-Order Execution”. In: *USENIX Security*. 2018. URL: <https://foreshadowattack.eu/foreshadow.pdf>.
- [2] Kit Murdock et al. “Plundervolt: Software-based Fault Injection Attacks against Intel SGX”. In: *Security & Privacy*. IEEE, 2020. URL: <https://plundervolt.com/doc/plundervolt.pdf>.
- [3] Ferdinand Brasser et al. “Software Grand Exposure: SGX Cache Attacks Are Practical”. In: *WOOT*. USENIX, 2017. URL: <https://www.usenix.org/system/files/conference/woot17/woot17-paper-brasser.pdf>.
- [4] *Confidential Computing concepts*. URL: <https://cloud.google.com/compute/confidential-vm/docs/about-cvm> (visited on 05/19/2022).
- [5] *Azure confidential computing documentation*. URL: <https://docs.microsoft.com/en-us/azure/confidential-computing/> (visited on 05/19/2022).

- [6] *AWS Nitro Enclaves*. URL: <https://aws.amazon.com/ec2/nitro/nitro-enclaves/> (visited on 04/15/2022).
- [7] Dayeol Lee et al. “Keystone: An Open Framework for Architecting Trusted Execution Environments”. In: *EuroSys*. ACM, 2020. URL: <https://dl.acm.org/doi/pdf/10.1145/3342195.3387532>.
- [8] *vsock(7) — Linux manual page*. URL: <https://man7.org/linux/man-pages/man7/vsock.7.html> (visited on 06/06/2022).
- [9] *kaniko*. URL: <https://github.com/GoogleContainerTools/kaniko> (visited on 04/20/2022).
- [10] Richard Barnes, Jacob Hoffman-Andrews, Daniel McCarney, and James Kasten. *RFC 8555: Automatic Certificate Management Environment (ACME)*. 2019. URL: <https://datatracker.ietf.org/doc/html/rfc8555> (visited on 06/03/2022).
- [11] *Challenge Types – Let’s Encrypt*. URL: <https://letsencrypt.org/docs/challenge-types/#http-01-challenge> (visited on 06/06/2022).
- [12] Jaehyuk Lee et al. “Hacking in Darkness: Return-oriented Programming against Secure Enclaves”. In: *USENIX Security*. 2017. URL: <https://www.usenix.org/system/files/conference/usenixsecurity17/sec17-lee-jaehyuk.pdf>.
- [13] Andrea Bittau et al. “PROCHLO: Strong Privacy for Analytics in the Crowd”. In: *SOSP*. ACM, 2017. URL: <https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/46411.pdf>.
- [14] Jun Xu, Jinliang Fan, Mostafa Ammar, and Sue B. Moon. “On the Design and Performance of Prefix-Preserving IP Traffic Trace Anonymization”. In: *Internet Measurement Workshop*. ACM, 2001. URL: <https://conferences.sigcomm.org/imc/2001/imw2001-papers/69.pdf>.
- [15] Ramakrishna Padmanabhan et al. “DynamIPs: Analyzing address assignment practices in IPv4 and IPv6”. In: *CoNEXT*. ACM, 2020. URL: https://www.prihter.com/dynamips_conext20.pdf.
- [16] *chi*. URL: <https://github.com/go-chi/chi/> (visited on 05/25/2022).
- [17] *Nitro Security Module Interface for Go*. URL: <https://github.com/hf/nsm> (visited on 05/25/2022).
- [18] *Linux networking in Golang*. URL: <https://github.com/milosgajdos/tenus> (visited on 05/25/2022).
- [19] *Baton*. URL: <https://github.com/americanexpress/baton> (visited on 04/18/2022).
- [20] Stefano Garzarella. *iperf3: A TCP, UDP, SCTP, and VSOCK network bandwidth measurement tool*. URL: <https://github.com/stefano-garzarella/iperf-vsock> (visited on 06/06/2022).
- [21] *The Underhanded C Contest*. 2015. URL: <http://www.underhanded-c.org> (visited on 04/18/2022).
- [22] Sergei Arnautov et al. “SCONE: Secure Linux Containers with Intel SGX”. In: *OSDI*. USENIX, 2016. URL: <https://www.usenix.org/system/files/conference/osdi16/osdi16-arnautov.pdf>.
- [23] Gordon King and Hans Wang. *HTTPa: HTTPS Attestable Protocol*. 2022. arXiv: 2110.07954v2 [cs.CR]. URL: <https://arxiv.org/pdf/2110.07954.pdf>.
- [24] Fan Zhang et al. “Town Crier: An Authenticated Data Feed for Smart Contracts”. In: *CCS*. ACM, 2016. URL: <https://dl.acm.org/doi/pdf/10.1145/2976749.2978326>.
- [25] Vikram Sharma Mailthody et al. “Safer Illinois and RokWall: Privacy Preserving University Health Apps for COVID-19”. In: *CoronaDef*. The Internet Society, 2021. URL: https://www.ndss-symposium.org/wp-content/uploads/coronadef2021_23001_paper.pdf.
- [26] Juhyeng Han, Seongmin Kim, Jaehyeong Ha, and Dongsu Han. “SGX-Box: Enabling Visibility on Encrypted Traffic using a Secure Middlebox Module”. In: *APNet*. ACM, 2017. URL: <https://conferences.sigcomm.org/events/apnet2017/papers/sgxbox-han.pdf>.
- [27] Moxie Marlinspike. *Technology preview: Private contact discovery for Signal*. Sept. 2017. URL: <https://signal.org/blog/private-contact-discovery/>.
- [28] Joshua Lund. *Technology Preview for secure value recovery*. Dec. 2019. URL: <https://signal.org/blog/secure-value-recovery/>.
- [29] *Asylo*. URL: <https://github.com/google/asylo> (visited on 05/16/2022).
- [30] *Open Enclave SDK*. URL: <https://github.com/openenclave/openenclave> (visited on 05/16/2022).
- [31] Huibo Wang et al. “Towards Memory Safe Enclave Programming with Rust-SGX”. In: *CCS*. ACM, 2019. URL: <https://dl.acm.org/doi/pdf/10.1145/3319535.3354241>.
- [32] Stephan van Schaik et al. “CacheOut: Leaking Data on Intel CPUs via Cache Evictions”. In: *Security & Privacy*. IEEE, 2021. URL: <https://cacheoutattack.com/files/CacheOut.pdf>.

- [33] Sangho Lee et al. “Inferring Fine-grained Control Flow Inside SGX Enclaves with Branch Shadowing”. In: *USENIX Security*. 2017. URL: <https://gts3.org/assets/papers/2017/lee:sgx-branch-shadow.pdf>.
- [34] Jo Van Bulck et al. “A Tale of Two Worlds: Assessing the Vulnerability of Enclave Shielding Runtimes”. In: *CCS*. ACM, 2019. URL: https://fahrplan.events.ccc.de/rc3/2020/Fahrplan/system/event_attachments/attachments/000/004/153/original/ccs19-tale.pdf.
- [35] Wenhao Wang et al. “Leaky Cauldron on the Dark Land: Understanding Memory Side-Channel Hazards in SGX”. In: *CCS*. ACM, 2017. URL: <https://donnod.github.io/files/papers/ccs17.pdf>.
- [36] Shufan Fei, Zheng Yan, Wenxiu Ding, and Haomeng Xie. “Security Vulnerabilities of SGX and Countermeasures: A Survey”. In: *ACM Computing Surveys* 54.6 (2021). URL: <https://dl.acm.org/doi/pdf/10.1145/3456631>.
- [37] Alexander Nilsson, Pegah Nikbakht Bideh, and Joakim Brorsson. *A Survey of Published Attacks on Intel SGX*. 2020. arXiv: 2006.13598 [cs.CR]. URL: <https://arxiv.org/pdf/2006.13598.pdf>.
- [38] Keegan Ryan. “Hardware-Backed Heist: Extracting ECDSA Keys from Qualcomm’s TrustZone”. In: *CCS*. ACM, 2019. URL: <https://sci-hub.se/https://doi.org/10.1145/3319535.3354197>.