Compiler-Assisted Semantic-Aware Encryption for Efficient and Secure Serverless Computing

Bongjun Kim®, Seonyeong Heo®, Student Member, IEEE, Jaeho Lee, Shinnung Jeong, Yongwoo Lee, and Hanjun Kim®, Member, IEEE

Abstract—Serverless computing like Function-as-a-Service (FaaS) is attractive for IoT service providers, liberating the providers from server maintenance. Since a data processing function is executed on the cloud instead of a dedicated server in the FaaS platform, the service users send their private data in their IoT devices to the third-party cloud, taking privacy leakage risks. Homomorphic encryption (HE) can preserve the privacy by enabling encrypted data processing on the cloud, but using HE for every data item incurs large computation and communication overheads. This work proposes SelectiveCrypt, a compiler-assisted semantic-aware encryption scheme that applies different cryptographic primitives depending on the operations on each data item. SelectiveCrypt homomorphically encrypts data items if arithmetic operations are applied to the data, while SelectiveCrypt encrypts data items with a symmetric key if the data are stored in the cloud without any arithmetic operation. The SelectiveCrypt framework consists of a compiler and its runtime system. The SelectiveCrypt compiler statically analyzes the data processing, determines an appropriate cryptographic primitive for each data item, and automatically transforms arithmetic operations into the homomorphic computation. The SelectiveCrypt runtime encrypts and decrypts the data items according to the static analysis result. This work evaluates the prototype SelectiveCrypt framework with five benchmarks that reflect real-world IoT scenarios. The evaluation results show that the SelectiveCrypt framework successfully reduces response time and communication overhead by 1.59 times and 9.61 times, respectively, compared with a HE scheme.

Index Terms—Cloud computing security, homomorphic encryption (HE), Internet of Things.

I. INTRODUCTION

SERVERLESS computing like Function-as-a-Service (FaaS) facilitates to implement scalable IoT services without maintaining a dedicated server. Popular IoT platforms [1]–[4] support FaaS services such as Amazon Web Services (AWS) Lambda, and Azure Functions. Using the FaaS services, users can load and execute a data processing function on the cloud on demand. Here, to offload data processing on the cloud, users should send plain IoT data to the cloud. For example, to execute a face recognition function on the cloud, users should send a camera image to the cloud.

Though FaaS liberates users from server maintenance, FaaS forces users to take privacy leakage risks on their IoT services. FaaS relies on third-party cloud platforms, and IoT devices send and store private user data to the untrusted cloud platforms. Since curious cloud service providers or malicious attackers may access to the plain IoT data on the cloud, the secure end-to-end communication is not enough for privacy preserving.

Previous work [5]–[10] has proposed encryption-based approaches to protect user data from the untrusted cloud. The work either uses symmetric-key encryption or homomorphic encryption (HE). Symmetric-key encryption offers relatively fast encryption and decryption, but it disregards the data processing in the cloud. On the other hand, HE enables encrypted data processing on the cloud, but it incurs huge computation and communication overheads. Thus, for users to securely execute a data processing function on the cloud, HE should be adopted in the FaaS service.

Although HE is unavoidable in the secure FaaS service, not all the data items require the HE. If a processing function in the FaaS service stores user data to the cloud storage just for archiving without any arithmetic operation, symmetric-key encryption is enough to protect the data. For example, if a processing function recognizes a known face with a camera image and archives a video clip without any data processing, the function only needs to encrypt the camera image.
in HE. Symmetric-key encryption is enough to safely store the video clip in the cloud. Therefore, depending on the semantics of the processing function, the FaaaS service can selectively encrypt user data either in HE or in symmetric-key encryption.

This work proposed SelectiveCrypt, a compiler-assisted semantic-aware encryption scheme that applies different cryptographic primitives, such as HE and symmetric-key encryption to data items according to their usage patterns in an FaaaS service. This work also implements a prototype SelectiveCrypt framework that automatically transforms a processing function in an FaaaS service into a secured function and supports encrypted data processing without modifying third-party cloud platforms. The framework consists of a compiler and its runtime system. The SelectiveCrypt compiler analyzes the function code and determines an appropriate cryptographic primitive for each data item. The compiler maps each data item into cryptographic primitives according to its semantic in the function and passes the mapping information to the SelectiveCrypt runtime. With the mapping information, the SelectiveCrypt runtime conducts encryption and decryption on data items. To support lightweight IoT devices, the SelectiveCrypt framework also allows the devices to offload encryption and decryption to the SelectiveCrypt runtime on a trusted proxy. Fig. 1 briefly illustrates the execution model of the SelectiveCrypt framework.

This work evaluates the prototype SelectiveCrypt framework with five benchmarks that reflect real-world IoT scenarios, such as face recognizer, batting estimator, gas leak detector, character recognizer, and sensor failure detector on the cloud. The evaluation results show that the SelectiveCrypt framework securely executes the benchmarks with 1.59 times shorter response time and 9.61 times less communication than a fully homomorphic encrypted scheme. In addition, with three microbenchmarks that implement representative IoT computations, such as average, matrix multiplication (MM), and dot product, this work deeply analyzes the performance of offloading encryption and decryption from a weak device to a powerful proxy.

The contributions of this work are as follows.
1) The SelectiveCrypt scheme that selectively and efficiently encrypts user data either in HE or in symmetric-key encryption depending on the semantics of a processing function in an FaaaS service.
2) The design of the SelectiveCrypt compiler that statically determines cryptographic primitives of data items and automatically transforms code to support encrypted data processing without modifying cloud platforms.
3) The design of the SelectiveCrypt runtime that provides seamless encryption and decryption across IoT devices and a trusted proxy.

II. BACKGROUND AND MOTIVATION

A. Function-as-a-Service in the IoT

FaaaS is an emerging type of cloud computing services. FaaaS allows users to execute code on the cloud without building and maintaining servers. With FaaaS services, users write, upload, and execute a function code on demand. Unlike private cloud servers, FaaaS services automatically manage multiple concurrent executions and charge according to how many or how long the code is executed. Therefore, the FaaaS service model can empower the IoT where the numbers of devices vary and events occur sporadically.

Existing IoT platforms [1]–[4] support FaaaS services, such as AWS Lambda, Azure Functions, Google Cloud Functions, and IBM Cloud Functions. The IoT platforms provide FaaaS SDKs for users to write a function that manipulates and stores IoT device data. To execute the function in the FaaaS services, the IoT platforms maintain connections with IoT devices and deliver messages from the IoT devices. In this way, the IoT platforms facilitate to implement IoT applications exploiting resources of the cloud. For example, an IoT application can use an FaaaS service for face recognition, which would be a huge computational burden for IoT devices with low computation power.

B. Homomorphic Encryption

HE, originally proposed by Rivest et al. [11], provides homomorphism between ciphertext and plaintext that guarantees the decryption result of computed ciphertext to coincide with the computed value from plaintext. Depending on what operations and how much operations can be performed on encrypted data, there are several classes of HE such as additive HE (AHE), leveled HE (LHE), and fully HE (FHE). AHE only allows addition on ciphertext and LHE allows a certain number of multiplications depending on the security parameters. The FHE scheme, proposed by Gentry [12], allows an
arbitrary number of operations to be performed on encrypted data, but its applicability is limited due to its high execution overheads. Subsequent FHE schemes, such as BGV [13], BFV [14], CKKS [15], and RNS-CKKS [16] have made significant efficiency improvement.

Among the FHE schemes, this work exploits the BFV scheme for HE. The security of the BFV scheme is based on the ring learning with errors (RLWE) problem [17]. By utilizing the RLWE problem, the BFV scheme maps plaintext on the ring \( R_T := \mathbb{Z}_T[x]/(x^N + 1) \) to ciphertext on the ring \( R_Q := \mathbb{Z}_Q[x]/(x^N + 1) \), where \( N, T \) and \( Q \) represent polynomial modulus degree, plaintext modulus, and ciphertext modulus, respectively. A user can guarantee a certain level of security by choosing these RLWE parameters. Since the RLWE parameters affect the multiplicative depth, encryption/decryption and computation time and security level, it is difficult for users without cryptography expertise to determine. With the advent of Microsoft SEAL [18], HElib [19], HEAAN [20], and PALISADE [21] that implement efficient FHE schemes including the BFV scheme as opensource libraries, the user can easily apply the FHE schemes to their applications. For \( N \) and \( T \) selected by the user at a given security level, the SEAL library that this work uses generates secret and public keys with the corresponding hard-coded \( Q \).

C. Privacy Preserving With Encryption

Since the IoT platforms force users to send their private IoT data to the cloud in FaaS, the users should take privacy leakage risks on their IoT services. Although the IoT platforms provide an end-to-end secure channel to defend the privacy leakage against man-in-the-middle attacks, the platforms decrypt the data in the cloud to execute the processing function. Thus, the data in the cloud can be leaked by curious cloud service providers or malicious attackers. In other words, the end-to-end secure communication cannot preserve privacy from the untrusted cloud.

To securely process and store the data in the cloud, previous work [5]–[10], [22]–[24] suggests encrypting data before sending to the cloud. Some works [5]–[8] propose systems that automatically encrypt data for secure query processing in the cloud database. P²-SWAN [9] proposes a mobile computing framework that provides seamless data encryption in the edge. There also exist commercial software solutions [10], [22]–[24] that encrypt data stored in the cloud. These encryption-based approaches either use symmetric-key encryption or HE. Symmetric-key encryption is less computationally expensive than HE, but symmetric-key encryption disregards data processing in the cloud. Thus, the applicability of symmetric-key encryption is limited to cloud storage services only. On the other hand, HE enables to process encrypted data in the cloud. However, HE is much more computationally expensive than symmetric-key encryption, and thus increases performance overheads of the encryption.

Fig. 2 shows the ciphertext size, encryption time, and decryption time of symmetric-key and HE for different plaintext sizes. This work uses AES and BFV algorithms for symmetric encryption and HE, respectively. To evaluate the two encryption algorithms, this work executes the algorithms on Raspberry Pi 3B+. On average, AES obtains 4.01 times smaller ciphertext size, 58.6 times shorter encryption time, and 11.7 times shorter decryption time than BFV. The results demonstrate that symmetric-key encryption incurs much less communication and computation overheads than HE.

Fortunately, the encryption overheads can be reduced because not all the data need HE in the FaaS service. HE is necessary only when the cloud processes data. If the cloud only archives data for logging or delivers data among devices as a bridge, HE is not necessary because there is no data processing in the cloud. Therefore, depending on the data usage patterns, the FaaS service can selectively apply different cryptographic primitives, reducing the HE overheads.

III. SYSTEM DESIGN

This work proposes an efficient semantic-aware encryption scheme for privacy preserving in the IoT, called SelectiveCrypt and its compiler-runtime prototype framework. The SelectiveCrypt compiler analyzes a function code, decides cryptographic primitives for each data item between HE and symmetric-key encryption, and automatically transforms the original code to support encrypted data processing. The SelectiveCrypt runtime encrypts a data item with the corresponding cryptographic primitive before sending to the cloud. Here, without any modification on the cloud, the cloud can process the transformed code with the encrypted data. This section describes a threat model of this work, the SelectiveCrypt compiler and the SelectiveCrypt runtime in detail.

A. Threat Model

The IoT environment that the SelectiveCrypt framework operates consists of three participants: 1) IoT device; 2) Proxy; and 3) cloud. An IoT device publishes data items to topics or subscribes data items from interesting topics. When a user tries to deploy a function code, the user manually chooses and authenticates IoT devices as legitimate participants. A trusted proxy is one of powerful IoT devices such as desktops and smartphones within the trusted network such as a home network. If there is no powerful nor trustworthy device, SelectiveCrypt only supports on-device encryption without offloading the encryption to the trusted proxy. Each IoT device and proxy runs a preinstalled SelectiveCrypt runtime that manages encryption and keys to protect application-level data. The cloud belongs to third-party cloud service providers (e.g., Amazon AWS, Microsoft Azure, and Google Cloud) and provides various services such as computation (e.g., AWS Lambda) and storage (e.g., AWS S3 and DynamoDB).

The cloud provides an authentication and authorization infrastructure to secure communication among participants and prevent man-in-the-middle attacks, but the cloud does not guarantee that it will not exploit user data beyond security boundaries. Existing work [25]–[34] points out security concerns about the privacy and integrity of outsourced data storage and processing, assuming the cloud providers or attackers as potential threats. Especially, several papers [31]–[34] consider
Algorithm 1: Coloring Algorithm

Input: A data dependence graph \( G = (V, E) \)

Output: A coloring table \( CT \) that contains the assigned colors of all nodes in \( V \)

\[
\text{while } V \text{ is not empty do}
\]

\[
\begin{align*}
&v \leftarrow \text{Draw a vertex in } V \\
&CC \leftarrow \{\}, Q \leftarrow \{v\} \\
&C \leftarrow \text{Color}_{\text{SYM}} \\
&\text{// Find connected components} \\
&\text{while } Q \text{ is not empty do} \\
&\quad v \leftarrow \text{Draw a vertex in } Q \\
&\quad CC \leftarrow CC \cup \{v\} \\
&\quad \text{if } v \text{ is an arithmetic operation then} \\
&\quad \quad C \leftarrow \text{Color}_{\text{HE}} \quad \quad \quad \text{// Find connected components} \\
&\quad \text{end} \\
&\quad \text{for } (v', v) \in \text{Incoming}[v] \text{ do} \\
&\quad \quad \text{if } v' \notin CC \text{ then} \\
&\quad \quad \quad \text{Insert } v' \text{ into } Q \\
&\quad \quad \text{end} \\
&\quad \text{end} \\
&\quad \text{for } (v, v') \in \text{Outgoing}[v] \text{ do} \\
&\quad \quad \text{if } v' \notin CC \text{ then} \\
&\quad \quad \quad \text{Insert } v' \text{ into } Q \\
&\quad \quad \text{end} \\
&\quad \text{end} \\
&\text{end} \\
&\text{// Update the color table} \\
&\text{for } v \in CC \text{ do} \\
&\quad CT[v] \leftarrow C \\
&\text{end} \\
&V \leftarrow V \setminus CC \\
\end{align*}
\]

the cloud server as honest-but-curious, which means that the cloud server honestly executes requested services, but it may intentionally not forget user data. For example, although users request to delete their data, the cloud may not delete the data for their financial purposes. In addition, the cloud may use user data without authorization to train their object detection models or behavior prediction models.

This work assumes that the SelectiveCrypt runtime is installed on IoT devices and the trusted proxy with integrity. The deploy manager performs secure deployment processes, including key distribution and FaaS function code deployment. This work also assumes that data are delivered exactly where it should be delivered (confidentiality) without data manipulation (data integrity). Since this work considers only the cloud as a potential threat, the malicious act of either IoT devices or the trusted proxy is out of the scope of this threat model. Moreover, since the underlying SEAL library does not support comparison operations for homomorphically encrypted data, this work targets only applications where outsourced data are not used on conditional branches, and its threat model excludes the timing-based side-channel attack such as guessing the private data by exploiting unbalanced branches.

Original Function Code
1: \text{image} = \text{get}('/camera/image')
2: \text{video} = \text{get}('/camera/video')
3: \text{result} = \text{recognize}() \quad \text{// SelectiveCrypt}
4: \text{DB.save}() \quad \text{// SelectiveCrypt}
5: \text{pub}('/recog/result', \text{result})

Configuration File
\begin{align*}
\text{target_func: [ } & \text{"face_recog"}, \\
& \text{receive_func: [ } & \text{"get"}, \\
& \text{...} & \text{]}} \\
\end{align*}

SelectiveCrypt Compiler

1) Dependence Analysis

\begin{align*}
\text{Data Dependence Graph} \\
\text{depends on} & \quad \text{data sink} \quad \text{data source} \\
\text{perform arithmetic operations} & \quad \text{execute function code} \\
\end{align*}

2) Marking

\begin{align*}
\text{Marked Data Dependence Graph} \\
\text{data sink} & \quad \text{data source} \\
\text{execute function code} & \quad \text{execute function code} \\
\end{align*}

3) Coloring

\begin{align*}
\text{Colored Data Dependence Graph} \\
\text{SYM} & \quad \text{HE} \\
\end{align*}

4) Code Transformation

\begin{align*}
\text{Transformed Function Code} \\
1: \text{image} = \text{CipherText()} \\
2: \text{video} = \text{get}('/camera/video') \\
3: \text{result} = \text{recognize}() \quad \text{// SelectiveCrypt} \quad \text{// SelectiveCrypt} \\
4: \text{DB.save}() \quad \text{// SelectiveCrypt} \quad \text{// SelectiveCrypt} \\
5: \text{pub}('/recog/result', \text{result}) \quad \text{// SelectiveCrypt} \\
\end{align*}

Fig. 3. Overall compilation process of SelectiveCrypt.

\begin{align*}
\text{Decryption Table} \\
\text{Configuration File} \\
\text{Original Function Code} \\
\end{align*}

(b)

(a)

Fig. 4. Parts of the face recognizer code example and its transformed code by SelectiveCrypt. \( nKey \) represents relinearization key. (a) Original code. (b) Transformed code.

B. SelectiveCrypt Compiler

The SelectiveCrypt compiler receives a function code that takes unencrypted inputs, and generates a function code that takes encrypted inputs. The compiler determines an appropriate cryptographic primitive for each data item in the function code, and transforms the original code to enable encrypted data processing. Note that the input function code contains computations that operate on plaintexts without any HE-specific instructions. Fig. 3 illustrates the overall compilation process for an example pseudocode. The example code implements a face recognition application for home security. When the
face recognition function is triggered, the function acquires an image and a video clip from an IP camera (instructions 1 and 2), recognizes a face in the image (instruction 3), archives the video clip to the database (instruction 4), and publishes the recognition result (instruction 5).

The compilation process consists of four major steps: 1) dependence analysis; 2) marking; 3) coloring; and 4) code transformation. The first step is analyzing data dependences on a function code. When an instruction uses the result of the other instruction, the compiler marks the two instructions dependent. With the dependence analysis results, the compiler builds a data dependence graph, where each vertex represents an instruction and each edge indicates the dependence between two instructions.

The second step is marking arithmetic instructions, data sources and data sinks in the data dependence graph. An arithmetic instruction is an instruction that performs arithmetic operations. A data source is an instruction that extracts a data item from an event or acquires a data item from an IoT device. In the example code, instruction 1 can be a data source because instruction 1 extracts the image data from the inputs. A data sink is an instruction that publishes a data item. In the example code, instruction 5 can be a data sink because instruction 5 publishes the face detection result. In general, the compiler can find data sources and data sinks in the function code based on the communication API that the cloud platform provides.

The third step, coloring, is assigning appropriate cryptographic primitives for each instruction with the marked data dependence graph. Algorithm 1 explains how the compiler colors instructions. For each instruction that performs arithmetic operations, the compiler finds its reachable data sources and sinks in the dependence graph, and colors all the relevant instructions with HE. Then, the compiler colors all the uncolored instructions, including the uncolored data sources and sinks as SYM (symmetric-key encryption).

The final step is transforming the original code to support encrypted data processing. For data annotated with HE in the colored data dependence graph, the compiler transforms their source instructions to receive encrypted data. For example, since the input image is colored with HE in Fig. 1, the compiler inserts an instruction that initializes the image object to contain ciphertext objects. Note that no transformation is required for the arithmetic operations between their sources and sinks because the Microsoft SEAL library [18], utilized by the SelectiveCrypt compiler as the backend, implements operator overloading for its arithmetic operations, such as multiplication, addition, subtraction, squaring, and exponentiation.

However, since the size of the ciphertext increases after each homomorphic multiplication, the compiler applies the HE-specific operation relinearize to the multiplication result immediately after each homomorphic multiplication. For example, Fig. 4 shows how the compiler transforms parts of the recognize function code in the face recognizer into HE enabled code. The compiler inserts the relinearize instruction immediately after the homomorphic multiplication like line 15 at Fig. 4(b).

Finally, the compiler adds their sinks in the description table to inform that the published data are in form of ciphertexts. If the function code invokes another function, the SelectiveCrypt compiler duplicates the original function, and recursively transforms the duplicated function in the same way. For example, in Fig. 1, the SelectiveCrypt compiler replaces the original recognize function with the recognize\textsuperscript{HE} method that supports the encrypted data processing. The transformed recognize\textsuperscript{HE} method will contain HE-specific instructions [line 15 at Fig. 4(b)].

As the compilation result, the compiler returns the transformed function code, an encryption table, and a decryption table. The encryption and decryption tables describe cryptographic primitives for each data item. Here, this work assumes that each data item has a unique data identifier, and the compiler can extract the data identifiers from the function code. Note that the format of data identifiers can differ across different IoT platforms. For example, for the IoT platforms that use the MQTT protocol, a data identifier can be an MQTT topic such as /camera/image.

C. SelectiveCrypt Runtime

The SelectiveCrypt runtime deploys public keys to IoT devices, encrypts data items before sending the data items to the cloud, and decrypts encrypted data items received from the cloud. Fig. 5 illustrates the overall process of the SelectiveCrypt runtime. The SelectiveCrypt runtime runs on IoT devices and trusted proxy. If an IoT device can have too weak computation power to support encryption and decryption, the SelectiveCrypt framework allows the device to offload encryption and decryption to a trusted proxy. In this case, the trusted proxy acts like a communication broker that collects plain data items from IoT devices, encrypts the data items, and publishes the encrypted data items to the cloud on behalf of the IoT devices.
The deploy manager in the SelectiveCrypt deploys compiled results in four steps. First, the deploy manager uploads the transformed function code to the cloud (step 1). Second, the deploy manager prompts the user to choose a master IoT device that will decrypt data items in the decryption table, and makes the IoT device generate keys for encryption and decryption (steps 2 and 3). Finally, the deploy manager distributes the encryption table, the decryption table, and the keys to IoT devices (step 4).

The deploy manager uses different key distribution mechanisms for symmetric-key encryption and HE. For symmetric-key encryption, the deploy manager distributes the same symmetric key to the IoT devices that will encrypt or decrypt data items with symmetric-key encryption. For HE, the deploy manager distributes the same public key to the devices that will encrypt data items with HE. In addition, the deploy manager delivers the public key to the cloud because the cloud also needs a public key to process the data items.

For the example in Fig. 5, the deploy manager requests the door lock to generate a pair of private and public keys. According to the request, the door lock generates keys and passes the public key to the deploy manager. The deploy manager distributes the public key for other devices to apply HE.

The deploy manager keeps a private key for decryption in an only one IoT device, called master device. Since only the master IoT device has a private key, the device can decrypt data items encrypted with HE, and the other devices including the cloud can only process data items without decryption. This work introduces the master runtime that mediates data encryption as an extension, and installs the runtime extension on the master device.

The SelectiveCrypt runtime automatically encrypts and decrypts data items according to the encryption and decryption tables. The online part of Fig. 5 illustrates how the runtime encrypts and decrypts data items for the example function in Fig. 3. Before sending a data item to the cloud, the runtime checks whether the data item is in the encryption table, and encrypts the data item using the prescribed cryptographic primitive. For example, after the runtime triggers the lambda function (step 4), the runtime on the device encrypts the data such as video and image on the device (steps 1 and 2).

When receiving a data item from the cloud, the runtime checks whether the data item is in the decryption table. If so, the runtime decrypts the data item using the prescribed cryptographic primitive. For example, the trusted proxy receives a data item which identifier is /recog/result, it finds the data item in the decryption table and decrypts the data item using a private key on behalf of a resource-constrained doorlock (Step 3). If the runtime has no key to decrypt the data item, the runtime reports an error that it fails to decrypt the data item. Here, while the SelectiveCrypt runtime triggers an event and receives a termination signal through the MQTT protocol, the runtime sends program input data and receives output data through the S3 storage service due to the payload size limitations of the MQTT protocol itself.

If an IoT device cannot afford SelectiveCrypt runtime, the device offloads its cryptographic task into the trusted proxy. For example, since the doorlock is too weak to perform cryptographic tasks, the deploy manager installs the decryption table in the trusted proxy and delegates decryption authority to the trusted proxy. On the other hand, when the device has strong computation power as the IP camera, its runtime encrypts the data such as video and image on the device.

IV. SECURITY ANALYSIS

This section proves that the SelectiveCrypt framework maintains the higher or equal level of security compared to the existing homomorphic-only encryption. Since the SelectiveCrypt framework encrypts parts of data using symmetric-key encryption rather than HE, the problem to prove is whether the semantic-aware hybrid encryption scheme can have the higher or equal security level compared to homomorphic-only encryption.

Table I shows notations used in this section. With given encryption parameters $N$, $Q$, and $T$, the HE scheme generates a key $hk$. Here, $N$ is polynomial modulus degree for polynomial modulus $x^N + 1$. $Q$ and $T$ are coefficient modulus and plaintext modulus, respectively. The symmetric encryption scheme generates a key $sk$ with a key size, $m$-bits. $E_{hk}(d_i)$ and $E_{sk}(d_i)$ stand for ciphertexts that the HE and symmetric encryption schemes encrypt the $i$th data $d_i$, respectively. $\lambda_{E_{hk}(d_i)}$ and $\lambda_{E_{sk}(d_i)}$ are security levels measured in bits. $\lambda_{E_{hk}(d_i)}$ and $\lambda_{E_{sk}(d_i)}$ mean that $2^{\lambda_{E_{hk}(d_i)}}$ and $2^{\lambda_{E_{sk}(d_i)}}$ operations are required to break the HE and symmetric-key encryption of the data $d_i$ encrypted with the secret keys $hk$ and $sk$, respectively.

**Problem Statement:** The semantic-aware hybrid encryption preserves a security level that is equal to or higher than the HE scheme.

**Assumptions:** For the security analysis, this work assumes the following.

1) **Security Parameters:** A user can guarantee a certain level of security by choosing encryption parameters, such as $N$, $Q$, and $T$ for HE scheme and $m$ for symmetric-key encryption scheme. For example, this work chooses $N$ as 8192 for polynomial modulus $x^N + 1$ and uses hard-coded default coefficient modulus $Q$ to guarantee 256-bits security level corresponding to AES-256-b security level. For symmetric encryption, by specifying a key size, $m$-bit, as 256 or higher bits, the following holds true:

$$\lambda_{E_{hk}(d_i)} \geq \lambda_{E_{sk}(d_i)}.$$  \hspace{1cm} (1)

2) **Security Key and Level of Multiple Cryptographic Primitives:** The security level of a system that uses multiple encryption schemes is the security level of a scheme that requires the least amount of operations to break it.

**Theorem 1:** For the adversary to attack the semantic-aware encrypted ciphertexts, the number of required operations is equal to or higher than the number of required operations to attack homomorphically encrypted ciphertexts

$$\min(\lambda_{E_{hk}(d_1)}, \lambda_{E_{hk}(d_2)}, \ldots, \lambda_{E_{hk}(d_n)})$$

$$\geq \min(\lambda_{E_{sk}(d_1)}, \lambda_{E_{sk}(d_2)}, \ldots, \lambda_{E_{sk}(d_n)})$$ \hspace{1cm} (2)

for $n$ data items. Here, $ck$ is either $hk$ or $sk$.  

Authorized licensed use limited to: Yonsei Univ. Downloaded on March 26, 2021 at 07:58:57 UTC from IEEE Xplore. Restrictions apply.
**Table I**

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N, Q$</td>
<td>Homomorphic encryption security parameters</td>
</tr>
<tr>
<td>$m$</td>
<td>Symmetric encryption key size</td>
</tr>
<tr>
<td>$h$</td>
<td>A key for homomorphic encryption</td>
</tr>
<tr>
<td>$s$</td>
<td>A key for symmetric encryption</td>
</tr>
<tr>
<td>$ck$</td>
<td>A key for homomorphic or symmetric encryption</td>
</tr>
<tr>
<td>$d_k$</td>
<td>$i$-th data item for outsourcing</td>
</tr>
<tr>
<td>$E_{hk}(d_i)$</td>
<td>Homomorphically encrypted $i$-th data item with key $hk$</td>
</tr>
<tr>
<td>$E_{sk}(d_i)$</td>
<td>Symmetrically encrypted $i$-th data item with key $sk$</td>
</tr>
<tr>
<td>$\lambda_{E_{hk}}(d_i)$</td>
<td>A security level of homomorphic encryption for $E_{hk}(d_i)$</td>
</tr>
<tr>
<td>$\lambda_{E_{sk}}(d_i)$</td>
<td>A security level of symmetric encryption for $E_{sk}(d_i)$</td>
</tr>
</tbody>
</table>

**Lemma 1:**

$$\lambda_{E_{ck}}(d_k) \geq \lambda_{E_{hk}}(d_k).$$

**Proof:**

Case $ck$ is $hk$: $\lambda_{E_{ck}}(d_k) = \lambda_{E_{hk}}(d_k)$.

Case $ck$ is $sk$: $\lambda_{E_{sk}}(d_k) = \lambda_{E_{sk}}(d_k) \geq \lambda_{E_{sk}}(d_k)$ from (1).

Thus, $\lambda_{E_{ck}}(d_k) \geq \lambda_{E_{sk}}(d_k)$.

**Proof:** We will prove Theorem 1 by induction that for $n$ data items, $\min(\lambda_{E_{ck}}(d_1), \ldots, \lambda_{E_{ck}}(d_n)) \geq \min(\lambda_{E_{hk}}(d_1), \ldots, \lambda_{E_{sk}}(d_n))$.

**Base Case:** When $n = 1$, $\min(\lambda_{E_{ck}}(d_1)) \geq \min(\lambda_{E_{sk}}(d_1))$.

**Induction Case:** If (2) is true for $k$ data items, (2) will be true for $k + 1$ data items (induction hypothesis).

To simplify the equation, let us introduce two variables, $E_{ck}$ and $E_{hk}$.

$$E_{ck} = \min(\lambda_{E_{ck}}(d_1), \ldots, \lambda_{E_{ck}}(d_k))$$

$$E_{hk} = \min(\lambda_{E_{hk}}(d_1), \ldots, \lambda_{E_{sk}}(d_k)).$$

From the induction hypothesis, $E_{ck} \geq E_{hk}$.

From Lemma 1, $\lambda_{E_{ck}}(d_{k+1}) \geq \lambda_{E_{sk}}(d_{k+1})$.

Case $E_{ck} \geq \lambda_{E_{ck}}(d_{k+1})$:

$$\min\left(\lambda_{E_{ck}}, \lambda_{E_{ck}}(d_{k+1})\right) = \lambda_{E_{ck}}(d_{k+1}) \geq \lambda_{E_{sk}}(d_{k+1}) \geq \min\left(\lambda_{E_{sk}}, \lambda_{E_{sk}}(d_{k+1})\right).$$

Case $\lambda_{E_{ck}}(d_{k+1}) \geq E_{ck}$:

$$\min\left(\lambda_{E_{ck}}, \lambda_{E_{ck}}(d_{k+1})\right) = \lambda_{E_{ck}} \geq \lambda_{E_{sk}} \geq \min\left(\lambda_{E_{sk}}, \lambda_{E_{sk}}(d_{k+1})\right).$$

From (4) and (5), the induction hypothesis is true.

Thus, Theorem 1 holds for all the positive integer $n$.

**V. EVALUATION**

This work implements and evaluates the prototype SelectiveCrypt framework with AWSs, which supports both IoT and FaaS services. This work uses the AES-256 implementation in the OpenSSL library for symmetric-key encryption and the BFV [14] implementations in Microsoft SEAL library [18] for HE and its python wrapper Pyhel library [35] and pyHeal library [36]. We have made a prototype implementation of SelectiveCrypt publicly available via GitHub [37].

To ensure the $\lambda$-bit security level in HE scheme, this work chooses the encryption parameters by referring to the recommendation parameter tables listed in [18] and [38]. However, encryption parameters affect not only the security level but also the number of multiplications that can be performed on a ciphertext, that is, multiplicative depth. Since the FaaS service is designed for tasks that are executed for a short time on requests and provides limited memory usage and short time-out, it is suitable for running relatively shallow artificial neural networks. Therefore, this work chooses fixed encryption parameters to be multiplicatively deep enough to execute the neural network defined in Microsoft CryptoNets [39] correctly. If the user wants to support arbitrary multiplicative depth, SelectiveCrypt can be extended to use bootstrapping to support the FHE scheme.

This work chooses 8192 for polynomial modulus degree $N$ where polynomial modulus is a form of $x^N + 1$. By choosing polynomial modulus degree $N$ and desired security level (256 b), SEAL library gives hard-coded default values for coefficient modulus $Q$ where $Q$ is 118 b in length. For plaintext modulus $T$, this work chooses 109951122689 where $T$ can be any positive integer at most 60 b in length.

To compare the computation efficiency of different computing models, this work incorporates three different IoT system participants: 1) a cloud platform; 2) a trusted proxy; and 3) IoT devices. This work uses various cloud platform services such as Amazon AWS Lambda, IoT Core, and S3 services, which provide function execution, communication management, and cloud storage, respectively. This work also uses a desktop server and an embedded board, Dell XPS 8700 (Intel Core i7-4770) and Odroid N2 (ARM Cortex-A53) as trusted proxies, and small compute boards, Raspberry Pi 3B+ as IoT devices.

**A. Benchmark Description**

To evaluate the SelectiveCrypt framework, this work implements five benchmarks that reflect real-world IoT scenarios.

1) **Privacy-Preserving Face Recognizer:** It receives an image and a video clip from an IoT device, uses CryptoNets [39] to perform inference, and saves the video clip to the cloud storage. CryptoNets is a deep neural network that operates on homomorphically encrypted data. The SelectiveCrypt encrypts the image with HE and the video clip with symmetric-key encryption.

2) **Battling Estimator in Sport:** It receives hitting data of a baseball player and runs a multilayer perceptron to predict next turn behavior. Perceptron is a single layer...
neural network that consists of one output layer and activation function. SelectiveCrypt encrypts the hitting data with HE while encrypting all the strategies for the player with symmetric-key encryption.

3) **Gas Leak Detector:** It receives encrypted data from gas sensors and uses a perceptron to validate gas leak. SelectiveCrypt encrypts sensor values with HE while encrypting their sensor position and nearby temperatures with symmetric-key.

4) **Hand-Written Character Recognizer:** It receives encrypted hand-written numbers on the check and returns digits using homomorphic logistic regression. In addition, the metadata of the user are encrypted by symmetric-key and saved in the cloud.

5) **Sensor Failure Detector:** It analyzes 2000 sensors using linear regression and notifies a failure to the owner if a sensor shows a strange behavior. SelectiveCrypt encrypts the sensor values with HE, and their metadata with symmetric-key encryption.

### B. Execution Time and Communication Amount

For the benchmarks, this work measures the execution time and communication amount of the IoT device for six different computing models. In order to analyze the impact of semantic-awareness for a single transaction, this work measures the execution time and communication amount of the IoT device by setting their batch size as one.

1) **HE on Device Without Proxy (HE-D):** An IoT device encrypts all data items with HE and directly sends the encrypted data items to the cloud. The cloud executes the benchmark with the data items and returns the encrypted result to the device. The IoT device decrypts the result.

2) **HE With Weak/Powerful Proxy (HE-WP/HE-PP):** An IoT device sends plain data items to the computationally weak or powerful proxy. The proxy encrypts all the data items with HE and sends the encrypted data items to the cloud. The cloud executes the benchmark with the data items and returns the encrypted result to the proxy. The proxy decrypts and sends the result to the IoT device.

3) **Semantic-Aware Encryption on Device Without Proxy (SE-D):** An IoT device encrypts data items with different cryptographic primitives according to the usage of the data items in the benchmark. Then, the IoT device directly sends the encrypted data items to the cloud. The cloud executes the benchmark with the data items and returns the encrypted result to the device. The IoT device decrypts the result.

4) **Semantic-Aware Encryption With Weak/Powerful Proxy (SE-WP/SE-PP):** An IoT device sends plain data items to the proxy. The proxy encrypts data items with different cryptographic primitives according to the usage of the data items and sends the encrypted data items to the cloud. The cloud executes the benchmark with the data items, and returns the encrypted result to the proxy. The proxy decrypts and sends the result to the IoT device.

Fig. 6 shows the execution time breakdown of the benchmarks for different computing models. This work partitions the
total execution time into initialization, device-proxy communication, encryption, cloud communication, and computation. Initialization includes the setup time of HE. Device-proxy communication means communication latency between an end device and the proxy. Encryption consists of homomorphic and symmetric-key encryption time. Cloud communication means communication latency between a device or the proxy and the cloud. Cloud computation means the computation cost to execute the benchmark on the cloud.

Fig. 6 shows that SelectiveCrypt reduces total execution time by 1.50 times and 1.59 times on geometric mean for SE-WP and SE-PP models compared with HE-D model respectively. SelectiveCrypt reduces encryption and communication latency by avoiding HE for storage only data items. Moreover, proxy computing further reduces the encryption time by offloading time-consuming encryption to the proxy.

SelectiveCrypt reduces encryption and communication latency by avoiding HE for storage only data items. Moreover, proxy computing further reduces the encryption time by offloading time-consuming encryption to the proxy.

Fig. 7 shows the communication amount breakdown of the benchmark execution for different computing models in perspective of the IoT device. The SE-D model obtains 9.61 times less communication amount on geometric mean for SE-WP and SE-PP models compared with HE-D model respectively. SelectiveCrypt reduces encryption and communication latency by avoiding HE for storage only data items. Moreover, proxy computing further reduces the encryption time by offloading time-consuming encryption to the proxy.

SelectiveCrypt reduces encryption and communication latency by avoiding HE for storage only data items. Moreover, proxy computing further reduces the encryption time by offloading time-consuming encryption to the proxy.

C. Sensitivity Analysis on Encryption Types

Since SelectiveCrypt reduces the response time by replacing HE into symmetric-key encryption, the ratio between homomorphic and symmetric-key encryption highly affects the overall performance. Fig. 8 shows that the encryption time and ciphertext size for different encryption ratios. As the ratio of symmetric-key encryption increases, the total encryption time and ciphertext size linearly decrease. Considering the ciphertext size affects the communication overhead, applying symmetric-key encryption as much as possible maximizes the performance improvement.

D. Performance Analysis on Proxy-Oriented Outsourcing

This work designs three microbenchmarks to evaluate how the trusted proxy can reduce computation overheads of encryption and decryption, considering the frequently used computations in IoT applications.

1) **Average (AV):** AV calculates the average for floating point numbers. AV is for evaluating the addition performance.

2) **Matrix Multiplication (MM):** MM conducts 2-D MM. MM is for evaluating the multiplication and batching performance.

3) **Dot Product (DOT):** DOT calculates the dot product of 2-D matrices.

For each microbenchmark, this work measures the execution time for two different computing models as follows.

1) **Device Encryption (D):** An IoT device encrypts all data items and directly sends the encrypted data items to the cloud. The cloud executes a microbenchmark with the data items and returns the encrypted result to the device. The IoT device decrypts the result.

2) **Proxy Encryption (P):** An IoT device sends plain data items to the proxy. The proxy encrypts the data items, and sends the encrypted data items to the cloud. The cloud executes a microbenchmark with the encrypted data and returns the encrypted result to the proxy. The proxy decrypts and sends the result to the IoT device.

Note that all the microbenchmarks consist of pure functional computations, and thus all data items are supposed to be encrypted with HE.

Fig. 9 shows the execution time breakdown of each microbenchmark for different element sizes. This work partitions the total execution time into the time for sending data items to the proxy (for the proxy encryption model only), the time for encoding and encrypting data items, the time for sending data items to the cloud, and the time for the other operations including the microbenchmark execution on the cloud. On average, the proxy encryption model obtains 1.6 times less total execution time than the device encryption model. Moreover, the proxy encryption model obtains 6.59 times less encoding and encryption time and 9.21 times less time for sending data items to the cloud. Therefore, even
TABLE II
COMPARISON OF ENCRYPTION-BASED APPROACHES

<table>
<thead>
<tr>
<th>System</th>
<th>Domain</th>
<th>Threat Model</th>
<th>Encryption Scheme</th>
<th>Data Manipulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES-based encryption gateway</td>
<td>General</td>
<td>Untrusted</td>
<td>Trusted</td>
<td>Proxy</td>
</tr>
<tr>
<td>Virshwanath et al. [25]</td>
<td>IoT</td>
<td>Untrusted</td>
<td>Proxy</td>
<td>AES</td>
</tr>
<tr>
<td>Secure cloud storage [26]–[34]</td>
<td>Query</td>
<td>Untrusted</td>
<td>Device or Proxy</td>
<td>Multiple</td>
</tr>
<tr>
<td>CryptDB [5]</td>
<td>Query</td>
<td>Untrusted</td>
<td>Device (Mobile)</td>
<td>Multiple</td>
</tr>
<tr>
<td>Talos [6], Poster [7], and Pilatus [8]</td>
<td>IoT</td>
<td>Untrusted</td>
<td>Device or Proxy</td>
<td>AES or PHE</td>
</tr>
<tr>
<td>P²-SWAN [9]</td>
<td>IoT</td>
<td>Untrusted</td>
<td>Trusted</td>
<td>AES and LHE</td>
</tr>
<tr>
<td>This work</td>
<td>IoT</td>
<td>Untrusted</td>
<td>Trusted</td>
<td>Possible</td>
</tr>
</tbody>
</table>

through the proxy encryption model requires extra communications between the IoT device and the proxy, the proxy encryption model benefits more from computation offloading to the proxy.

E. Compilation Time

This work measures the compilation time of the SelectiveCrypt compiler for the benchmarks. Since the final output of the SelectiveCrypt compiler is not a binary but a python code that runs on the AWS Lambda, so the compiler is basically a source-to-source translator. The compiler analyzes and transforms an original code at an abstract syntax tree (AST) level, and then emits higher level readable python codes. The measurement results show that the SelectiveCrypt compiler takes an average of 16.655 ms and up to 25.898 ms for the five benchmarks.

VI. RELATED WORK

Privacy Preserving With Encryption: Previous work [5]–[10], [22]–[34] has applied encryption schemes for preserving privacy in various domains. Table II summarizes related work of this paper.

P²-SWAN [9] is a mobile computing framework to securely perform edge computation in third-party untrusted data centers. By applying the Paillier HE [40] to the sensor data, P²-SWAN achieves real-time encryption and encrypted computation on both mobile devices and edge cloud platforms. However, P²-SWAN applies the same cryptographic primitive for all sensor data. On the other hand, this work applies different cryptographic primitives according to how data item is used in a function. Therefore, this work can more efficiently provide privacy preserving with encryption.

CryptDB [5] provides encrypted query processing with a trusted database proxy that encrypts and decrypts data while intercepting communications. CryptDB applies different encryption schemes to data depending on SQL queries. Introducing a new cryptographic primitive for join operations, CryptDB can protect data from compromised database or administrators without harming the semantics of queries. Due to large computational cost, CryptDB avoids HE for all and uses HEs for specific operations only.

Extending CryptDB [5], Shafagh et al. [6], [7] propose IoT systems to prevent data breach from database compromise and network-based attacks. For resource-constrained IoT devices, Talos [6] and Poster [7] optimize order-preserving and HE algorithms by tailoring the algorithms for integers. To exploit near-user computation with edge computing [41], [42], Talos and Poster allow IoT devices to encrypt and decrypt their own data without a trusted proxy, which offers more privacy.

Pilatus [8] further extends Talos [6] to enable secure data sharing between different users. With Pilatus, a user generates a token for the cloud to re-encrypt data to share the data with another user. Pilatus allows users to revoke their keys to terminate data sharing. Compared with CryptDB [5] and Talos [6], Pilatus largely reduces the decryption time by optimizing the partial HE algorithm with Chinese Remainder Theorem. The previous work on CryptDB [5]–[7] targets SQL query processing, whereas this work considers general IoT applications.

There exist other works [25]–[34] that propose secure cloud storage based on user-side encryption. Since data in the cloud are encrypted by users, the cloud cannot read the content of the data without decryption keys. Under the assumption of the “honest but curious” cloud server model, these work let users control their private data through their own key management schemes. Commercial software solutions [22]–[24] are also available for seamless data encryption on the broker using symmetric-key encryption. However, they disregard data processing in the cloud and mainly support cloud storage services only, limiting their applicability.

Confidential Cloud Computing: Cloud providers offer confidential cloud computing [43]–[47] based on trusted execution environments (TEEs) such as Intel SGX [48]. Using TEEs, confidential cloud computing protects privacy data from privileged accesses of malicious insiders, host OSes, and hypervisors. Whereas confidential cloud computing requires special security hardware modules, the SelectiveCrypt framework enables private function execution using encryption without any security hardware module.

Information Flow Analysis: Information flow analysis is to detect security violation when sensitive data propagate through program execution. JFlow [49] is a language extension that permits static checking of information flow annotations for privacy preserving. Extending JFlow [50], Zheng et al. [51] applied the language-based techniques for distributed systems. Their work enables secure program partitioning when unreliable hosts exist by replicating computation and data. Other work [52] proposes a compiler that automatically partitions a program with locality annotations into distributed programs. To
protection variable updates among distributed programs, the compiler inserts cryptographic operations. With information flow analysis, this work can extend the SelectiveCrypt compiler to perform more precise information flow analysis of privacy data by introducing extra annotations at the language level.

**General-Purpose Compilers for Privacy-Preserving Computation Using HE:** Previous work [53]–[56] proposes general-purpose compilers for privacy-preserving compilation with various optimizations using different HE schemes. Cingulata [53] compiles C++ programs to BFV [14] and TFHE [57] scheme-based privacy-preserving implementation. ALCHEMY [54] compiles DSL on Haskell using BGV [13] scheme-based Haskell library as a backend. ALCHEMY selects proper parameters for the HE library using its statically typed system. RAMPARTS [55] compiles a Julia program into PALISADE [21]-based implementation applying common optimizations, such as subexpression elimination and constant folding. EVA [56] proposes its own intermediate representation and compiles a program into SEAL [18]-based implementation. EVA optimizes arithmetic circuits based on graph rewriting rules, finds optimal parameters, and automatically inserts HE maintenance operations, such as relinearization and rescaling. The existing compilers proposed in previous work, however, are not aware of domain knowledge to selectively apply different cryptographic schemes depending on data usage patterns.

VII. CONCLUSION

This work proposed a new compiler-assisted semantic-aware encryption scheme named SelectiveCrypt. SelectiveCrypt applies different cryptographic primitives to each data item depending on their operations, thus allowing efficient and secure computation in FaaS services. By offloading encryption and decryption from a lightweight IoT device to a powerful trusted proxy, SelectiveCrypt can further reduce the encryption overheads. The evaluation with five benchmarks showed that SelectiveCrypt successfully reduces computation and communication overheads compared with an HE only scheme.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers and editors for their insightful comments and suggestions and the CoreLab for their support and feedback during this work.

REFERENCES


Bongjun Kim received the B.S. degree in computer science and engineering from Pohang University of Science and Technology, Pohang, South Korea, in 2016, where he is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering.

He is currently working on a compiler framework for many-core neural processing units. His research focuses on compiler optimization and security enhancement for the Internet of Things framework.

Seonyeong Heo (Student Member, IEEE) received the B.S. degree in computer science and engineering from Pohang University of Science and Technology, Pohang, South Korea, in 2016, where she is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering.

Her research interests include real-time system design for deep neural networks and the Internet of Things.

Jaeho Lee received the B.S. degree in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2020, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering.

His research focuses on system security for storage devices and high-level synthesis optimization.

Shinming Jeong received the B.S. degree in creative IT engineering from Pohang University of Science and Technology, Pohang, South Korea, in 2019. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, Yonsei University, Seoul, South Korea.

Her research interests are compiler framework for many-core neural processing unit and parallel programming in compiler.

Yongwoo Lee received the B.S. degree in computer science and engineering from POSTECH, Pohang, South Korea, in 2019. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, Yonsei University, Seoul, South Korea.

His research focuses on compiler and runtime systems for homomorphic encryption.

Hanjun Kim (Member, IEEE) received the B.S. degree in electrical engineering from Seoul National University, Seoul, South Korea in 2007, and the M.A. and Ph.D. degree in computer science from Princeton University, Princeton, NJ, USA, in 2009 and 2013, respectively.

He is an Associate Professor with the Department of Electrical and Electronic Engineering, Yonsei University, Seoul. From 2013 to 2018, he was an Assistant and Associate Professor with the Departments of Creative IT Engineering and Computer Science and Engineering, POSTECH, Pohang, South Korea. His research interests include computer architecture and compiler optimization for distributed and emerging systems.

Dr. Kim was awarded the Intel Corporation Ph.D. Fellowship and the Siebel Scholarship in 2012.