ABSTRACT

There are a number of domain-specific programming languages for secure computation. Out of those, the ones that are based on generic programming languages support mixing different protocol primitives and enable implementing a wider, possibly more efficient range of protocols. On the one hand, this may result in better protocol performance. On the other hand, this may lead to insecure protocols. In this paper we present a security type system that enables mixing protocol primitives in a generic programming language, but also ensures that well-typed programs are secure in the semi-honest model. Consequently, a compiled protocol must be secure. We show an extension of the L1 language with our security type system and evaluate the implementation of two protocols from the literature. This shows that our security type system supports the provably secure implementation even of complex protocols.

Categories and Subject Descriptors

D.4.6 [Operating Systems]: Security and Protection—Cryptographic controls; D.3.4 [Programming Languages]: Processors—Compilers

General Terms

Security, Programming Languages

Keywords

Secure Two-Party Computation, Type System, Programming, Domain-Specific Language, Information Flow

1. INTRODUCTION

Secure (two-party) computation [37] allows two parties to compute a function \( f \) over their joint, private inputs \( x \) and \( y \), respectively. No party can infer anything about the other party's input (e.g. \( y \)) except what can be inferred from one's own input (e.g. \( x \)) and output (e.g. \( f(x, y) \)). Secure computation has many applications, e.g. in the financial sector, and has been successfully deployed in commercial and industrial settings [5, 6, 7, 23].

Secure computation protocols are notoriously difficult to implement. First, they can encompass arbitrary functionality in the joint computation. Second, they need to follow a rigorous approach to security. Often special protocols – mixing several different primitives – are developed for important problems. This is expected to provide more efficient protocols due to insight into the problem domain [16]. These protocols then require a manual verification and security proof.

Current domain-specific programming languages for secure computation [2, 4, 9, 18, 28, 35] do not adequately address these problems. Either they are tied to a specific protocol [2, 4, 18, 28] or they enable implementing insecure protocols [9, 19, 35]. On the one hand, if they are tied to a specific protocol, then this protocol may be proven secure manually independent of the functionality. Such a proof extends to all protocols implemented in this language, but the language prevents implementing many special, possibly more efficient protocols. On the other hand, if the programming language is built upon a generic programming language, such as Python [9] or Java [19, 35], all special protocols can be implemented. Yet, this allows the programmer to also implement insecure protocols that do not withstand security verification.

In this paper we address this problem in a novel way. We build upon the L1 language [35] which is a domain-specific language extension for secure computation based on Java that allows mixing several different protocol primitives. Although we use this specific language, our approach is generic and can be adapted for any domain-specific language that allows mixing protocols and is not tied to a protocol, e.g. [9, 19]. We augment the L1 language with a novel security type system. This type system provably ensures that well-typed programs are secure in the semi-honest model of secure computation. Loosely speaking, in the semi-honest model – as defined by Goldreich [15] – all parties follow the protocol description, but may keep a record of the interaction and try to infer additional information about the other party’s input. Protocols secure in the semi-honest model provably prevent any such inference which includes many real-world attacks, such as insider attacks.

Through the use of the type system the compiler statically verifies the security of the protocol during compilation. Only secure protocols are compiled. The programmer is immediately notified about any (possible) security violation.
We evaluate our security type system for the L1 language by implementing two examples. The first one is a simple multiplication protocol based on [14]. The second one is a complex protocol for privacy-preserving string processing from [21]. This example shows that even such protocols – using a wide variety of protocol primitives in intricate ways – can be effectively implemented in our type system. Furthermore, the security proof of our type system also extends to this new implementation. This underpins the protocol’s manual security proof from [21] by formal verification.

In summary, this paper contributes

- a security type system for mixed protocol secure computation.
- an integration of this type system into the L1 language.
- a proof that any well-typed program is secure in the semi-honest model.
- an evaluation using two exemplary protocols for multiplication and substring creation. This shows that implementing any functionality and even complex protocols using our type system is feasible.

The remainder of this paper is structured as follows. In Section 2 we review the related work in secure computation and type systems. Section 3 reviews the protocol primitives implemented for secure computation and Section 4 reviews the L1 language. Sections 5 and 6 present our main results. Section 5 describes how re-randomization is implemented and Section 6 describes the security type system. We show our examples in Section 7 and conclude the paper in Section 8.

2. RELATED WORK

Our work is related to domain-specific languages (and extensions) for secure computation [2, 4, 9, 18, 19, 28, 32, 35] and type systems for information flow security [12, 29, 34, 36].

Domain-specific languages (DSL) for secure computations can be coarsely classified into those based on a specific secure computation protocol, such as [2, 4, 18, 28], and those that extend a generic programming language [9, 19, 35]. DSLs based on specific secure computation protocols implement functionality solely based on those protocols. This may not limit implementable functionality, since these protocols are generic and support any functionality, but it may limit performance, since problem-specific protocol primitives are usually expected to yield better efficiency. Nevertheless, once the specific protocols is proven secure, any implemented functionality is also secure.

The first such domain-specific language for secure computation was FairPlay [28]. It implemented Yao’s two-party, garbled circuit protocol [37]. It was later extended to multi-party computations in [2] based on the multi-party version of Yao’s protocol [1]. It introduced the SFDL programming language which provided an abstraction of the ideal functionality, i.e. the function computed by the secure protocol. Programs in SFDL are translated into a binary circuit which is interpreted using Yao’s protocol. Secure multi-party computations were also implemented by ShareMind [4] based on the information-theoretically secure protocol of [3]. Its programming language is rather simplistic and resembles machine language. This makes programming harder, but also makes it easier to ensure security. All these single primitive DSLs generally suffer from poor performance of the compiled protocols. Mixing primitives in a DSL, such as Yao’s protocol and homomorphic encryption, was introduced in TASTY [18]. Nevertheless, they restricted themselves to one provably secure protocol of [25] and enabled annotations in the language naming the primitive to use.

DSLs as extensions of existing, generic programming languages can implement any protocol based on mixing (almost) arbitrary primitives. This allows implementing the most efficient protocols, but also does not prevent the programmer from making mistakes and implementing insecure protocols. The first such language is VIFF [9] based on the Python language. Although it is catered for the multi-party protocol using linear secret sharing of [8] which is also considered as the basis for the related SCML programming language of [32], in its implementation it supports the full power of Python. Similar extensions have been presented for the Java language [19, 35]. The framework of [19] is based on Yao’s protocol [37] and currently produces the most efficient protocols, but provides very little language support. A significant portion of its efficiency gain is based on optimizations by the programmer, which also invites mistakes likely at the expense of security. Implementing some such optimizations in the compiler was proposed in [22].

Our security type system is based on the L1 language [35] for mixed protocol secure computation, although it can be adapted for other mixed-protocol languages. This language also currently provides little support to the programmer for protocol security.

Our idea of supporting security in a programming language by a type system has been applied to information flow security. Information flow is concerned with confidentiality breaches between principals in computer systems [26]. Its lattice policy model of security levels has been introduced in [10]. Later, it was shown that compliance with such a policy can be statically verified given the program source code [11]. Volpano et al. were the first to recognize that these policies can also be implemented using a type system [36]. Their type system also influenced our construction of our secure computation security type system.

Since type systems can be effectively handled by programmers as well as compilers, this sparked recent interest in language-based information flow security. A very good survey of current research can be found in [34]. There is also an implementation of a type system for information flow security in Java called JIF [29].

Recently information flow types have been applied to homomorphic encryption [12]. These do not yet cover secure computations using homomorphic encryption, since all secure computations involve admissible information flows. Therefore the basic typing assumption of non-interference does not hold in secure computation. Our security type system must therefore cater for a more complicated use case.

3. SECURE COMPUTATION

We implement secure computations using the primitives of homomorphic encryption, secret sharing, garbled circuits and oblivious transfer. These primitives can be combined in many ways and even potentially result in an insecure protocol. For simplicity we only consider secure two-party computation.
3.1 Homomorphic Encryption

Secure computation can be implemented based on additively homomorphic encryption. On the one hand, opposed to fully homomorphic encryption [13] additively homomorphic encryption only implements addition (modulo a key-dependent constant) as the homomorphic operation. On the other hand, additively homomorphic encryption is as fast as standard public-key cryptography.

Let $E(x)$ denote the encryption of plaintext $x$ encrypted under $X$’s (Alice’s or Bob’s) public key and $D(x)$ the corresponding ciphertext of the encryption. Then the homomorphism can be expressed as

$$D_x(E_x(x)) = x + y$$

The following property can be easily derived

$$D_x(E_x(x)^y) = xy$$

In our implementation we mainly use Paillier’s encryption system [33], although our language also supports different encryption systems such as Naccache-Stern [31] and allows the programmer to extend the available encryption systems. Paillier’s and Naccache-Stern’s encryption systems are not only public-key, but also secure against chosen plaintext attacks (IND-CPA). This semantic security implies that two different ciphertexts – even of the same plaintext – cannot be distinguished (without the private key).

In secure computation based on homomorphic encryption it is usually sufficient to use one key per party. Our type system - differently than [12] - therefore does not protect against mixing ciphertexts from different keys.

3.2 Secret Sharing

In order to implement the full functionality of secure computation we augment homomorphic encryption with secret sharing. Each variable is secretly shared between Alice and Bob. Let $x$ be the variable secretly shared and $p$ be the modulus of the homomorphic encryption. Then Alice has $x_A$ and Bob has $x_B$, such that

$$x = x_A + x_B \mod p$$

In the following we show how secret sharing and homomorphic encryption can be used to implement any functionality in secure computation. Many other combinations of homomorphic encryption and secret sharing may already be insecure, e.g. by revealing dependent shares. In order to implement any ideal functionality it suffices to implement addition and multiplication. Addition of $x = x_A + x_B$ and $y = y_A + y_B$ (of the same bit-length $l$) can be implemented locally by addition of each party’s shares. Multiplication $z = x \cdot y$ needs to be implemented as a protocol. Let $r$ be a uniformly random number in $Z_p$. We use the commonly used protocol in Figure 1 from [14],

\[
A \rightarrow B \quad E_A(x_A), E_A(y_A) \\
B \rightarrow A \quad E_A(c) = E_A(x_A)^y \cdot E_A(y_A)^x \cdot E_A(r) \\
A \quad z_A = x_A y_A + c \mod p \\
B \quad z_B = x_B y_B \mod p
\]

Figure 1: Multiplication Protocol

It is easy to verify that $z_A + z_B = (x_A + x_B)(y_A + y_B)$.

3.3 Oblivious Transfer

Oblivious transfer (OT) is a protocol between a sender and a receiver. As input the sender has $n$ messages $x_0, \ldots, x_{n-1}$ and the receiver has an index $i$ ($0 \leq i < n$). After the protocol the receiver obtains $x_i$ as output. A secure OT protocol ensures that the sender has not learnt $i$ and the receiver has learnt nothing about the other messages $x_j$ ($j \neq i$).

OT can be used as a primitive in mixed secure computations. It facilitates a distributed “if” statement, where one party has the results of the branches and the other party the condition. Furthermore, although we do not implement this protocol in our language, OT can also be used as the sole primitive to implement generic secure computation protocols [15].

There are a number of efficient implementations of OT including [30]. Nevertheless, all OT involve some public key operations. In [20] the authors describe a technique to extend a constant number of OTs to polynomially many OTs using symmetric cryptography.

3.4 Garbled Circuits

Yao introduced secure two-party computation in [37]. He uses a technique called garbled circuits. Garbled circuits allow the computation of any function, but in our language we combine garbled circuits with secret shares and somewhat restrict its functionality. Our restriction is that each input and output (function result) must be a secret share. Let Alice’s input be $a$ and Bob’s input $b$ which both may be a secret share. Let $f$ be the function to be computed using garbled circuits. We then automatically augment the circuit using an additional input $r$ by the circuit encryptor. The circuit then computes the function $f'$:

$$f'(a, b, r) = f(a, b) - r$$

The circuit encryptor obtains no output from the secure computation, but uses its locally chosen $r$ as return value. The circuit decryptor obtains the new function result $f'$. These two return values are secret shares of the function result $f(a, b)$ which can be subsequently used in computations using homomorphic encryptions or garbled circuits.

Secrets as inputs are automatically reconstructed at the beginning of the function. For example, let $x_A, y_A$ and $x_B, y_B$ be shares of $x$ and $y$, respectively. Then the function $f(x, y) = x < y$ to compare $x$ and $y$ is implemented as follows:

$$f'(x_A, y_A, x_B, y_B, r) = (x_A + x_B < (y_A + y_B)) - r$$

4. THE L1 LANGUAGE

We extend the L1 language [35] for mixed-protocol secure computation with our security type system. L1 is a simplification and extension of Java that incorporates features to better support secure computation. The two L1 features important for this paper are messaging support and player-specific code.

Messaging support enables the parties to send each other messages. It is implemented using a framework for asynchronous communication available to all parties. In its basic form the send command is invoked as in Listing 1. Here, we send the contents of the variable $\texttt{var}$ to the player with identifier 2 under the message name $\texttt{var\_id}$. The variable $\texttt{var}$ is cleared after the send command, such that it cannot be reused. This ensures that each variable is sent at most once (which is also always sufficient).

\[
A \rightarrow B \quad E_A(x_A), E_A(y_A) \\
B \rightarrow A \quad E_A(c) = E_A(x_A)^y \cdot E_A(y_A)^x \cdot E_A(r) \\
A \quad z_A = x_A y_A + c \mod p \\
B \quad z_B = x_B y_B \mod p
\]
Player identifiers are assigned globally in a configuration file at the start of the protocol execution. An identifier is associated with an IP address and port number where the player’s code receives messages. Message names must be unique for a receiving party. Otherwise, messages may be lost due to overlapping communication.

The other party can receive a previously sent value using the read command. We show in Listing 2 an example where the player receives content under the message name var_id.

```
Listing 2: Receive Command
A particular design principle of the L1 language is that the code for all players is integrated into one program. This one (and the same) program is then executed by all players. This corresponds to secure computation where the functionality (and the protocol) is known to all players in advance. Now in order to implement the sending and receiving of messages between the parties – using the send and read commands – we use player-specific code.

Player-specific code is only executed at a player with a specific identifier. This player’s identifier is placed at the start of the statement. If no such identifier is present, the statement is executed by all parties. We can therefore implement the send and receive commands within the same program. The following Listing 3 combines the code of the previous two listings.

```

Abstraction of a secret share, but converting a secret share to an integer requires a type cast. When an expression is assigned to a secret share, a modulo operation is performed automatically. We declare a secret share at each party as in Listing 4.

```
Listing 4: Secret Share Declaration
There are several ciphertext data types. It not only depends on the encryption system used, but also for each public key encryption system there are two types of ciphertexts: one where the player has the private key and one where the player has only the public key. Our assumption is that for ciphertexts where the player has the private key no other player can access the plaintext, i.e. the private key is indeed confidential. Similar to secret shares we parameterize the ciphertext type with the respective (public or private) key.

A ciphertext contains by definition a secret share as payload. This is not explicitly declared in our language extension. Furthermore, this secret share has the same modulus as the homomorphic operation in the encryption system using the specified key. A ciphertext can be internally represented as an object containing multi-precision integers – the ciphertext value – and a reference to a key. Again, an integer can be implicitly converted into a ciphertext, but converting a ciphertext to an integer requires a type cast. We declare ciphertexts as in Listing 5.

```

4.2 Cryptographic Protocols
As cryptographic protocol primitives we need to support garbled circuits and oblivious transfer. We implement garbled circuits using two commands. Yao’s protocol has a sender who encrypts (and garbles) the circuit and a receiver who – after obtaining the decryption keys via oblivious transfer – decrypts the circuit. The ideal functionality to be computed in the circuit is specified as a function (in C like syntax). Each input variable is shared between both parties and reconstructed at the beginning of the circuit. Each party then inputs its shares as parameters to the command. The output is also automatically shared. The random variable r (using the modulus of the result) is automatically appended as input by the encryptor to the circuit. It is filled with a random input chosen by the command and returned as a result from the command. The output (the result of the circuit) is returned as a function result to the decryptor (subtracted by the random value of the encryptor).

In the ideal functionality we currently only support arithmetic operations (addition, subtraction, multiplication, division) and comparison (equal, lower, lower or equal). In Listing 6 we implement the circuit for the above function

```
Listing 6: Example Circuit
Oblivious transfer is also implemented via two commands: one for sending and one for receiving. The sending command takes an array as parameter. Its length is the number

```

4.1 Data Types
The L1 language has its own type system for basic, non-security relevant types. This type system is orthogonal to our security type system. L1 provides data types for multi-precision integers (int), public (pubk) and private (prvk) keys. We extend this for the purpose of our type system with data types for secret shares and ciphertexts.

L1 implements copy semantics for variable assignments. Even if a complex variable (such as a share, ciphertext or a multi-precision integer) is assigned, the assigned value is copied onto the memory of the assigned variable. This differs from the pointer semantics of Java where only a reference is assigned.

A secret share type is parameterized by the modulus of the share. We create only additive shares, i.e., let p be the modulus and s (s < p) be the secret. Then there are n (n ≥ 2) shares s_i, such that

```

A secret share can be internally represented as an object containing a multi-precision integer – the value of the secret share – and a reference to a multi-precision integer – the modulus. An integer can be implicitly converted into
of possible choices in the oblivious transfer. The input array is cleared after sending. The receiving command takes the index as a parameter and outputs the selected message. We need to specify the communicating party in the receive command as well, since it is an interactive protocol. In Listing 7 we show an example for an 1-out-of-2 oblivious transfer.\footnote{The star in line 1 is explained after introducing our security type system.\footnote{http://perldoc.perl.org/perlsec.html#Taint-mode}}

```
1 share<1 << 32> x, y;
2 share<2> r;
3
4 1: r = gcencrypt(2, "bool compare(int x, int y) {
5      "compare = x < y;
6    }", x, y);
7 2: r = gedecrypt(1, "compare", x, y);
```

Listing 6: Garbled Circuit Protocol

We define the semantics of an untainted variable as we use it in our security type system: An untainted variable is independently random, immutable, and use-once. Independently random implies that it contains independently randomly distributed data. It is immutable, as it can only be set by special commands and the randomness is chosen by the system. An untainted variable can only be used once and is cleared after use in a send or re-randomization command. We summarize these properties in Definition 1. Next, we need to describe how to create an untainted variable.

**Definition 1.** An untainted variable is an immutable object that contains random data independent of all other untainted variables. It can be used in send and re-randomization command and is cleared after use.

### 5. REVERSE TAILNT CHECKING

Our type system ensure that only freshly and independently chosen random variables are revealed to the other party. We distinguish these variables by typing them as untainted (versus tainted for variables during processing). Each basic data type of secret share, ciphertext or array can be tainted or untainted. We perform the operation of re-randomization for reverse taint checking in order to create the untainted type of variables.

Taint checking\footnote{http://perldoc.perl.org/perlsec.html#Taint-mode} is a programming language operation that allows to track whether inputs are processed safely. It helps to prevent certain common programming errors, such as SQL injections or buffer overflows. In taint checking mode each variable set by an input is tainted. Then after certain checks have been performed, e.g. in PERL using regular expressions, the variable can be cleared as untainted. Every time a tainted variable is used in a dangerous operation, such as accessing a database, the program is aborted with a fatal error.

We call our operation reverse taint checking, because on the one hand we use a similar variable tracking mechanism (although statically verifiable), but on the other hand the taint flag is set by local operations and checked on sending data. In detail our algorithm works as follows. Each variable is either tainted or untainted. All variables assigned by expressions are tainted, since they are result of a local computation. It does not matter whether the assignment expression is a read command, some arithmetic operation or even a constant. Every regularly assigned variable is tainted.

The fundamental check is that the sending commands do not accept tainted variables as parameters. If the parameter is tainted, then the program does not type check. Only untainted variables may be sent to the other parties.

```
1 1: share<p> \_x * [2];
2 2: share<p> y;
3 2: int b; y;
4
5 1: otsend(2, "id", \_x);
6 2: y = othead(1, "id", b);
```

Listing 7: Oblivious Transfer

Our different re-randomization operations are as follows:

- **Secret share**: A secret share is associated with a modulus p. We choose a fresh random number r uniformly in \(Z_p\). Let s be the value of the secret share. We return the updated

\[
\_s \leftarrow s + r \pmod{p}
\]

and copy unto the re-randomization parameter (given to the function) the new secret share with the value r.

- **Ciphertext with public key only**: The ciphertext and the contained secret share are associated with the same modulus p. We again choose a fresh random number r uniformly in \(Z_p\). We encrypt the negation of this random number r using the associated public key. Let \(E(-r)\) denote this encryption and c denote the ciphertext of the re-randomized variable. We return the following, updated ciphertext

\[
\_c \leftarrow c \cdot E(-r)
\]

and copy unto the re-randomization parameter the new secret share with the value r.

- **Ciphertext with private key**: We assume that the possessor of the private key is its sole possessor. Therefore the contained secret share cannot be accessed by any other party and there is no need to re-randomize it. Nevertheless, we need to re-randomize the ciphertext. We use the usual re-randomization of IND-CPA

```
1 share<p> s;
2 cipher<pubKey> c;
3 cipher<privKey> d;
4 share<p> x [2];
5 share<p> r;
6
7 \_s = rerandomize(s, r);
8 \_c = rerandomize(c, r);
9 \_d = rerandomize(d, null);
10 \_x = rerandomize(x, r);
```

Listing 8: Re-Randomization

```
secure, homomorphic encryption. Let $d$ denote this ciphertext. We return it as

$$\mathcal{A} \leftarrow d \cdot E(0)$$

The randomization parameter is not used.

- **Array**: For arrays we choose the randomization parameter (i.e. $r$, $E(r)$ or $E(0)$) only once. The same randomization is applied to each element of the array and the randomization parameter is copied as usual. The return array contains the re-randomized elements. Note that only the array is untainted by this operation. Each element in the array remains tainted. This enables using the array in an oblivious transfer command, but prevents the elements from being sent individually using the send command.

### 5.2 Re-Randomization with Untainted Variables

Sometimes it is necessary to choose random shares before performing an operation and sending the result. An instance of this requirement occurs in our second example of substring creation in Section 7.2. Usually an untainted variable is created during a re-randomization command and then sent over the network, but instead of directly sending it, the untainted variable can be re-randomized another variable. The semantics of this re-randomization operation is that the untainted input parameter is cleared after the re-randomization command (just as it is cleared after a send command), such that it cannot be reused. Yet, the return variable is then subsequently untainted and can be used as such. Listing 9 shows an example where an untainted variable is used to re-randomize another variable. The variable $s$ is null after the execution of this code.

The complete re-randomization function is depicted in Algorithm 1. It is easy to verify from Algorithm 1 that after re-randomization one variable -- the untainted one -- is independently randomly distributed as required by Definition 1. This holds even in case of re-randomization with untainted variables.

#### Listing 9: Re-Randomization with Untainted Variable

```java
share<p> r;
// s = s - r
s = rerandomize(s, r);
// s is used and then cleared
_t = rerandomize(t, _s);
```

### 5.3 Output

If all variables were only to be sent as re-randomized variables, no useful result could ever be obtained, i.e., every (useful) secure computation has an admissible information flow. Therefore we need to allow reconstruction of the output values from the secret shares. Let $s$ be a secret share designated as output. We use an output statement in L1 in order to declare output, but before we can output the result the parties need to exchange result secret shares. The output command is restricted to secret shares as parameter. They can only send the secret share if the variable is untainted. The programmer needs to declare that this variable is designated for output before the output is actually performed. We perform a mutual exchange of (result) shares between two parties as in Listing 10.

#### Listing 10: Share Reconstruction and Output

```java
share<p> s_prime;
s = output(s);
send(id()) % 2 + 1, s, "result_share");
s_prime = read("result_share");
s = s + s_prime;
output("result = ", s);
```

### 6. SECURITY TYPE SYSTEM

In our L1 language we augment the regular type system with a security type system where every expression carries both a type (such as secret share, ciphertext or array) and a security type (tainted or untainted). Following the definition of [34] our security type system is a collection of typing rules that describe what security type is assigned to a program (or expression), based on the types of sub-programs (sub-expressions). We write $\vdash \text{exp} : t$ to mean that the expression $\text{exp}$ has security type $t \in \{\text{tainted, untainted}\}$ according to the typing rules. This assertion is known as a typing judgment. Similar to the construction of the security
type system for information flow [36] we use a security context \([sc]\) associated with a label of the program counter. We use the security context to prevent the programmer from sending messages depending on the truth value of expressions with tainted variables. Untainted variables are random and therefore not useful for branching conditions. This limitation is important, since not only the content of a message may reveal information, but just sending the message itself. The typing judgment \([sc] \vdash C\) means that the program \(C\) is typable in the security context \(sc\).

Figure 2 presents the typing rules for a simplification of the L1 language. We omit loops, since they can only have a constant number of iterations in secure computation and can therefore be unrolled. We also omit the typing rules for the regular type system. Expression security types and security contexts can be either tainted or untainted. A typing rule is an inference rule that describes how types are assigned. The statements above the line must be fulfilled for the rule to be applied, yielding the statement below the line.

According to the rules \([E1-3]\), some variables \((u)\) have security type untainted. All other expressions and variables have type tainted. This includes expression which include (but are not limited to) occurrences of untainted variables. Rules \([R1-2]\) ensure that untainted variables can only be assigned by re-randomization and output statements, but are otherwise static.

Consider the rules \([M1-6]\). The message sending commands can only be executed in an untainted security context. Of those the commands that transmit plaintext messages as payload \((\text{send, otsend})\) can only transmit untainted variables. Received variables are always immediately tainted.

The typing rules \([C1-5]\) control the security context in a composite program. The commands \(\text{skip}\) and variable assignment are typable in any context. Branches must be typable in a tainted context. This is justified by the above requirement that message sending commands may not be executed depending on tainted variables. The rules \([C4-5]\) enable composing programs similar to [36] including the subsumption rule. It enables sending messages before or after a tainted context (branching).

In L1 we declare untainted variables using an asterisk after the type declaration. We show examples for a share and an untainted array in Listing 11 (and already in Listing 7).

6.1 Proofs

Goldreich [15] defines security in the semi-honest model. The view \(\text{VIEW}^\Pi(x, y)\) of a party during protocol \(\Pi\) on this party’s input \(x\) and the other party’s input \(y\) is its input \(x\), the outcome of its coin tosses and the messages received during the execution of the protocol.

**Definition 2.** We say a protocol \(\Pi\) computing \(f(x, y)\) is secure in the semi-honest model, if for each party there exist a polynomial-time simulator \(S\) given the party’s input and output that is computationally indistinguishable from the party’s view \(\text{VIEW}^\Pi(x, y)\):

\[
S(x, f(x, y)) \simeq \text{VIEW}^\Pi(x, y)
\]

Our main theorem states that well-typed program is compiled into a protocol secure in the semi-honest model.

\[
\vdash u : \text{untainted} \quad (E - 1) \\
\vdash \text{var} \neq u \quad (E - 2) \\
\vdash \text{exp} \neq u \\
\vdash \text{var} : \text{tainted} \quad (E - 3) \\
\vdash \text{var}_1 : \text{untainted} \\
\vdash \text{var} = \text{rerandomize}(\text{exp}, \text{var}_2) \quad (R - 1) \\
\vdash \text{var} : \text{untainted} \quad (R - 2) \\
\vdash \text{var} = \text{output}(\text{exp}) \quad (M - 1) \\
\vdash \text{var} : \text{tainted} \quad (M - 2) \\
\vdash \text{var} = \text{read}() \quad (M - 3) \\
\vdash \text{var} : \text{untainted} \quad (M - 4) \\
\vdash \text{var} = \text{outread}(\text{exp}) \quad (M - 5) \\
\vdash \text{var} : \text{tainted} \quad (M - 6) \\
\vdash \text{var} = \text{gcencrypt}(\text{exp}, \ldots) \\
\vdash \text{var} = \text{gcdecrypt}(\text{exp}, \ldots)
\]

\[
\vdash \text{skip} \quad (C - 1) \\
\vdash \text{var} : \text{tainted} \\
\vdash \text{var} = \text{exp} \quad (C - 2) \\
\vdash \text{if} \text{exp} \text{then} \text{C}_1 \text{else} \text{C}_2 \quad (C - 3) \\
\vdash \text{C}_1; \text{C}_2 \quad (C - 4) \\
\vdash \text{C}_1 ; \text{C}_2 \quad (C - 5)
\]

**Theorem 3.** Let program \(C\) implement function \(F\). Then if \(C\) is well-typed \((\vdash C)\), it implements \(F\) securely in the semi-honest model.

**Proof Outline:** We need to show that we can construct a simulator for each party’s view. The problem is that the view is determined by the messages received and not sent (where type safety applies). We therefore assume that the entire protocol is written in L1 using player-specific code. Then we simulate the messages received from the other party. We give only the construction for one party’s simulator – due to the symmetry of the language it also applies to the other party.

Furthermore, differently from information flow type system there is always an admissible information flow in secure computation. This is due to the nature of the joint computation, i.e. some result is revealed. It is therefore not possible to strictly distinguish between the program and the
information conveyed. We capture this using the output variables.

Proof. We now construct the simulator for one party. The simulator consists of input, coin tosses and messages received. Input can be simulated by the real input (which is given to the simulator) and coin tosses are chosen by the Java random number generator. This leaves the received messages to be simulated.

Let \( m_1, \ldots, m_n \) be the messages received. We then construct the simulator step-by-step, i.e. for each message received we add a simulated message. This means, we start with the empty simulator \( S_0 \) which does nothing. Then we construct \( S_i \) from \( S_{i-1} \) by appending a simulation of message \( m_i \) to \( S_{i-1} \). Therefore it is first necessary to note that the number of messages is constant.

Lemma 4. In a well-typed program \( \langle \cdot \rangle \mathcal{C} \) the number of sending statements \( (\text{send, olsend, gcencrypt}) \) is constant.

Note that no sending statement may appear inside a branching statement. We can complete the proof by induction over the structure of \( \mathcal{C} \).

We now construct a simulated message for each message \( m_i \) received. We refer to Goldreich’s composition theorem [15] for our cryptographic protocols. Oblivious transfer and Yao’s protocol can be simulated as an oracle transmitting only the resulting messages, as long as the implementations have been proven secure in the semi-honest model. The security proof for our implementation of oblivious transfer can be found in [30]. The security proof for generic Yao’s protocol can be found in [27]. We can therefore simulate only the resulting messages.

There are two types of messages received: output and intermediate.

Output messages: These are messages marked as output by the other party. Let \( n_i \) be a message received. In a correctly implemented program these messages are indeed output at the local party (after combination with a local share). Therefore the local party can simulate the message as follows: Let \( o \) be the output which is given to the simulator and \( l \) the local share for reconstructing this value.

The local share \( l \) can be computed from the local state of the party up to that point in the program. This must be the case, since otherwise the party could not produce the output. Then the message is simulated as

\[ m_i = o - l \quad (\text{mod} \ p) \]

Intermediate messages: These are messages re-randomized by the other party. Let \( n_i \) be such a message received. In a well-typed program these messages are all randomly independently distributed.

Lemma 5. In a well-typed program all re-randomized variables are independently randomly distributed.

This follows directly from our construction of re-randomization. After the algorithm of re-randomization (see Algorithm 1) only one untainted variable is assigned. This variable is always independently randomly distributed. Therefore the set of re-randomized variables is always a set of independently randomly distributed variables. Note that local and output variables may be dependent on the re-randomized variables.

The received message \( m \) is then simulated as an independent random variable depending on the type of the message (secret share or public-key ciphertext)

\[ m_i \leftarrow \mu \mathbb{Z}_p/E(\mathbb{Z}_p) \]

This completes the simulator – all messages can be simulated – and consequently the proof of semi-honest security.

7. EXAMPLES

7.1 Multiplication

We first show a simple example. We implement the multiplication protocol using homomorphic encryption and secret shares from Figure 1 based on [14]. This protocol shows that using our typed language any functionality can be implemented securely (Listing 12).

The encrypt command (lines 33-34) returns a tainted variable. Subsequently (lines 35-36) we re-randomize the ciphertext (with known private key). This may seem superfluous, since the randomness is fresh during encryption, but the contained share may be tainted in ciphertexts with public key only. We chose this re-randomization approach over a polymorphic command based on return type. We see in line 47 that using the rerandomize command we can omit explicitly adding a random variable. Instead we use the randomness introduced by the re-randomization command.

7.2 Substring Creation

As a second example we show the substring creation protocol of Jensen and Kerschbaum [21]. This protocol is one of a set of three for privacy-preserving string processing. It is particularly suited to show the power of our language and type system. First, it uses all of our primitives: Yao’s garbled circuits, oblivious transfer, secret shares and homomorphic encryption.

The protocol operates on strings. A string consists of a length field and a fixed-length array of characters. Each string is secretly shared between Alice and Bob, such that neither Alice nor Bob know its contents or even length.

The purpose of the protocol is to create a substring of length \( t \) of a string \( x \) starting at position \( t \). The resulting substring will also be secretly shared – just as the inputs, i.e. there is no output reconstruction. The protocol starts with a rotation of the shared string to the left by \( s \) characters. Then the trailing characters beyond the length field are masked, such that they are all set to 0 again. Each share of the string is masked separately, such that the masking protocol is executed twice with the roles reversed. The composed main protocol is shown in Listing 13.

The types for strings (containing length field and array) are defined in a separate header (not shown for brevity) This header of the protocol also loads the cryptographic keys for the homomorphic encryption. Both a private, public key-pair and a public key (of the other party) are read from disk.

In this protocol we use Naccache-Stern encryption [31], because the modulus of the homomorphic operation must be the same at both parties. This cannot be achieved using Paillier’s encryption system [33]. We use a fixed modulus of 256 in order to simplify integration with Yao’s protocol.

In [21] the authors use Goldwasser-Micali encryption [17],
1 int multiply (int s, int t) {
2   // key variables
3     1: prvk prvkey;
4     2: pubk pubkey;
5     3: // read keys from hard disk
6       1: {
7         2: prvkey = readKey("prv.key");
8         3: pubkey = getPublicKey(prvKey);
9       } 2: pubKey = readKey("pub.key");
10     4: int n = getModulus(pubKey);
11     5: share<n> x, y, z;
12     6: // initialize shares
13       1: x = s;
14       2: y = t;
15       3: // declare ciphertexts
16         1: {
17             2: cipher<prvKey> e, f, g;
18             3: cipher<prvKey>e, f, g;
19             4: share<n> r;
20         } 2: cipher<pubKey> e, f, g;
21     7: share<n> r;
22     8: // send 1st message
23       1: {
24           2: e = encrypt(privKey, x);
25           3: f = encrypt(privKey, y);
26           4: e = re-randomize(e, null);
27           5: f = re-randomize(f, null);
28           6: send(2, e, "E\text{A}(x,A)"");
29           7: send(2, f, "E\text{A}(y,A)"");
30       } 2: // receive 1st message, send 2nd
31       1: {
32           2: e = read("E\text{A}(x,A)");
33           3: f = read("E\text{A}(y,A)"");
34           4: g = modPow(e, y, n);
35           5: g = modPow(f, x, n);
36           6: g = re-randomize(g, r);
37           7: send(1, g, "E\text{A}(c)"");
38       } 2: // receive 2nd message, compute result
39       1: {
40           2: g = read("E\text{A}(c)";
41           3: z = decrypt(g) + x * y;
42           4: // also compute result at Bob
43           5: z = r + x * y;
44       } 2: return (int) z;
45
46 Listing 12: Secure Multiplication

1 string substring (string x, string y)
2     1: share<n> s,
3          2: share<n> t) {
4     3: string a, b, sub;
5     4: sub = rotate(x, sub);
6     5: sub.len = t;
7     6: 1: a = mask_send(sub);
8     7: a = mask_recv(sub);
9     8: b = mask_send(sub);
10    9: return add(a, b);
11}
12
13 Listing 13: Substring: Composed Protocol

1 string rotate (share<n> s, string in) {
2     1: string x, r;
3     2: string priv e;
4     3: _string pub y;
5     4: _string pub y;
6     5: x = in;
7     6: 1: x = rerandomize_shift(x, s);
8     7: x = rerandomize_shift(x, s);
9     8: _e = encrypt_string(x);
10    9: _e = encrypt_string(x);
11    10: send_string("1", _e);
12   11: } 2:
13      1: y = recv_string_pub("1");
14      2: y = recv_string_pub("1");
15      3: rotate_left(y, s);
16      4: rotate_right(r, s);
17      5: x = add(x, r);
18      6: send_string("2", _y);
19      7: send_string("2", _y);
20     8: _e = encrypt_string(x);
21     9: _e = encrypt_string(x);
22    10: send_string("3", _e);
23    11: } 2:
24       1: e = recv_string_priv("2");
25       2: e = recv_string_priv("2");
26       3: x = decrypt_string(e);
27       4: x = decrypt_string(e);
28       5: rotate_left(x, s);
29       6: rotate_left(x, s);
30       7: _x = rerandomize_string(x, _x);
31       8: _x = rerandomize_string(x, _x);
32       9: send_string("4", _y);
33      10: send_string("4", _y);
34     11: } 2:
35        1: e = recv_string_priv("4");
36        2: e = recv_string_priv("4");
37        3: x = decrypt_string(e);
38        4: x = decrypt_string(e);
39        5: rotate_left(x, s);
40        6: rotate_left(x, s);
41    12: return x;
42}
43
44 Listing 14: Substring: Rotation

1 string rerandomize_shift (string x, share<n> s) {
2     1: _string _x;
3     2: _string _x;
4     3: string r;
5     4: 5: x = rotate_right(x, s);
6     6: _x = rerandomize_string(x, r);
7     7: x = r;
8     8: x = rotate_left(x, s);
9     9: return _x;
10    10:
11}
12
13 Listing 15: Substring: Initial Re-randomization

because they need its properties in a subsequent protocol. In this paper we do not implement this subsequent protocol and can therefore use the more efficient Naccache-Stern encryption system. There are also some helper functions for adding (secret shares), rotating strings, re-randomization functions for the different string types, en-/decryption functions and message sending functions – all not shown for brevity.

Next we consider the sub-protocols of the substring protocol. The complete rotation protocol is presented in Listing 14. The rotation protocol rotates the string, such that the initial character is at position 0. This is achieved us-
ing the following technique: Each share is encrypted and sent to the other parties which then rotates it by its share of the initial position. It then re-randomizes it and returns it before it is rotated by the local share of the initial position. This requires a particular interlocking technique. Before Alice can send her share to Bob she needs to re-randomize it with the re-randomization values for Bob’s share (these will cancel out). She cannot do this after she receives the returned ciphertexts from Bob, since they are then already rotated. Therefore, in order to implement this protocol Alice needs to choose the re-randomization parameters for Bob’s share before she sends her share. This is achieved by re-randomization using untainted variables created in the rerandomize_shift function (Listing 15). This function creates an untainted variable containing a random share that is already rotated by the local share of the initial position.

We have implemented the entire protocol in one function where each party’s code is implemented using player-specific code. We chose this type of implementation, since the two sides of the protocol are quite asymmetric. In line 11 (Listing 14) there is the call of the initial re-randomization creating the untainted variable. This untainted variable is then used for re-randomization in line 31 (Listing 14) completing the interlock technique. Bob can simply add his re-randomization to his share in line 20 (Listing 14).

```c
1 string mask_send (string x) {
2   share<2> b;
3   string r, q;
4   string_pub y, e, f[2];
5   string_pub f[*2];
6   y = recv_string("5");
7   f[0].len = x.len;
8   f[1].len = x.len;
9   for (int i = 0; i < x.len; i++)
10      f[i].chars[i] =
11      encrypt(pubKeyExt, 1);
12      for (int i = x.len; i < n; i++)
13         f[i].chars[i] =
14         y.chars[i - x.len];
15      for (int i = 0; i < n - x.len; i++)
16         f[i].chars[i] =
17         y.chars[i + n - x.len];
18      for (int i = n - x.len; i < n - x.len; i++)
19         f[i].chars[i] =
20         encrypt(pubKeyExt, 0);
21      b = gcdecrypt("
22         bool compare(int x, int y) {
23            compare = x < y;
24      (int) x.len, n);
25      if (b == 1) {
26         e = f[0];
27         f[0] = f[1];
28         f[1] = e;
29      } f = rerandomize(f, r);
30      otsend("f");
31      q = recv_string("7");
32      q = add(q, r);
33   return q;
34 }
```

Listing 16: Substring: Masking Sender

The masking protocol creates a 0, 1 encrypted string which is used to mask the local share. Care has to be taken, if the shares of the length of the substring wrap around the modulus. Two cases need to be prepared: one in which they do and one in which they do not wrap around. The correct one is chosen by oblivious transfer. The condition is computed by a comparison implemented as a Yao’s protocol. Therefore we see also these primitives in this protocol.

The masking protocol is executed twice with roles reversed. Therefore we implemented each side of the protocol as a separate function. Listing 16 shows one side and Listing 17 the other. Then each side is called with the local input in the composed protocol (Listing 13 lines 6–9).

8. CONCLUSIONS

In this paper we consider the problem of ensuring semi-honest security of secure computations implemented in domain-specific languages. Currently, either the language is based on a specific protocol which has been proven secure or on a generic language that enables also implementing insecure protocols. We are the first to propose a type system to limit the implementable protocols to only secure ones, but also enable the programmer to freely choose the protocol primitives. We prove that any well-typed program is secure in the semi-honest model of secure computation. This presents a new trade-off between security and performance for the implementation of secure computations. Our type system enables implementing complex protocols, such as one of our examples for privacy-preserving string processing.
Future work is to extend the security guarantees to stronger security models, such as the malicious model. In the malicious model the parties may behave arbitrarily and still privacy of inputs and correctness of outputs are preserved. Such a security guarantee can be of practical relevance in highly sensitive secure computations such as e-voting.

9. REFERENCES


