Viaduct
An Extensible, Optimizing Compiler for Secure Distributed Programs

Coşku Acay∗
Cornell University
Ithaca, NY, USA
cacay@cs.cornell.edu

Rolph Recto∗
Cornell University
Ithaca, NY, USA
rr729@cornell.edu

Joshua Gancher
Cornell University
Ithaca, NY, USA
jrg358@cornell.edu

Andrew C. Myers
Cornell University
Ithaca, NY, USA
andru@cs.cornell.edu

Elaine Shi
Cornell University
Ithaca, NY, USA
runting@gmail.com

Abstract
Modern distributed systems involve interactions between principals with limited trust, so cryptographic mechanisms are needed to protect confidentiality and integrity. At the same time, most developers lack the training to securely employ cryptography. We present Viaduct, a compiler that transforms high-level programs into secure, efficient distributed realizations. Viaduct’s source language allows developers to declaratively specify security policies by annotating their programs with information flow labels. The compiler uses these labels to synthesize distributed programs that use cryptography efficiently while still defending the source-level security policy. The Viaduct approach is general, and can be easily extended with new security mechanisms.

Our implementation of the Viaduct compiler comes with an extensible runtime system that includes plug-in support for multiparty computation, commitments, and zero-knowledge proofs. We have evaluated the system on a set of benchmarks, and the results indicate that our approach is feasible and can use cryptography in efficient, nontrivial ways.

CCS Concepts:
• Security and privacy → Information flow control; Cryptography; Domain-specific security and privacy architectures.

Keywords: information flow, multiparty computation, zero knowledge

∗Equal contribution.

ACM Reference Format:

1 Introduction
Modern distributed applications such as federated systems and decentralized blockchains typically involve parties from multiple administrative domains each with its own security policy. Companies might be required by law (such as the European Union’s GDPR [21]) to protect user privacy when they process user data or share it with other companies. The lack of full trust among parties makes it difficult to develop such systems, especially when the security requirements necessitate the use of cryptographic mechanisms. Recent efforts from the cryptography community have pushed these mechanisms from theory to practical deployment [5], but a gap remains: they still require too much expertise to use successfully [15, 17, 22].

We introduce Viaduct, a system that makes it easier for non-expert programmers to develop secure distributed programs that employ cryptography. It puts a variety of sophisticated cryptographic mechanisms in the hands of developers, including secure multiparty computation (MPC) protocols, zero-knowledge proofs (ZKP), and commitment schemes. Viaduct’s security-typed language allows developers to annotate programs with information-flow labels to specify fine-grained security policies regarding the confidentiality and integrity of data and computation. An inference algorithm allows these annotations to be lightweight, and enables Viaduct to reject inherently insecure programs. Viaduct then enforces these policies by compiling high-level source code to secure distributed programs, automatically choosing efficient use of cryptography without sacrificing security. The compiler supports a range of cryptographic protocols whose security guarantees are characterized using information-flow...
labels. New protocols can be added to Viaduct by specifying their security properties and by implementing well-defined interfaces.

Although prior efforts have attempted to bridge this gap, most existing work focuses on compiling programs to a fixed set of cryptographic mechanisms. For example, some focus on compiling programs to MPC (e.g., Wysteria [40], ObliVM [33], SCALE-MAMBA [3]); others focus on ZKP (e.g., Pinocchio [37], Buffet [46], xjSNARK [31]). To our knowledge, by providing a unified abstraction to both specify security policies of programs and to specify security guarantees of cryptographic mechanisms, Viaduct is the first system to compile secure, distributed programs with an extensible suite of cryptography.

We make the following contributions:

- An algorithm to infer minimum consistent security requirements of data storage and computation for programs written in a security-typed language. (§3)
- A technique to compile secure distributed programs, deploying an extensible set of cryptographic protocols while minimizing a customizable notion of cost. (§4)
- An extensible runtime system for running compiled programs. Cryptographic mechanisms are added as plug-ins to the runtime. (§5, §6)
- An evaluation that shows that the Viaduct compiler can synthesize a wide variety of secure and efficient distributed programs, that the compilation technique is scalable, and that the annotation burden of the source language is minimal. (§7)
- An open-source implementation of the Viaduct compiler and runtime system.\(^1\)

\(^1\)Available at https://github.com/apl-cornell/viaduct.

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{architecture.png}
\caption{Architecture of Viaduct.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{example.png}
\caption{Implementation of the historical millionaires’ problem in Viaduct. Viaduct uses MPC for the comparison \(a < b\), but computes the minima locally.}
\end{figure}

2 Overview of Viaduct

Figure 1 gives a high-level overview of Viaduct. Its compiler takes a high-level source program partially annotated with information-flow labels. The compiler infers labels consistent with programmer-supplied annotations to determine security requirements for all program components. Then for each component the compiler selects a protocol that matches these requirements, guiding the selection with a cost model. The output is a secure and efficient distributed program, which hosts execute using the Viaduct runtime system. The Viaduct architecture has a small set of well-defined extension points, allowing developers to add support for new protocols with relative ease.

We give two examples to motivate and describe the Viaduct compilation process.

**Historical Millionaires’ Problem.** Our first example is a slightly modified version of the “millionaires’ problem” [47]. As in the classic formulation, two individuals, Alice and Bob, want to determine who has more money without revealing how much money they have to the other person. Rather than comparing their current wealth, in our “historical” variant Alice and Bob want to see who was richer at their poorest point, comparing their current wealth, in our “historical” variant.

Figure 2 shows an implementation of the historical millionaires’ problem in Viaduct. The program compares Alice’s lowest wealth with Bob’s, and outputs the answer (\(b_{\text{richer}}\)) to both Alice and Bob.

Viaduct programs must specify the hosts that participate in the program, along with the authority that each host has, as shown in lines 1–2. All security policies in Viaduct are represented using security labels (in blue), which are defined formally in §2.1. Security labels capture both confidentiality and integrity. For example, host alice is given label \(A \land B^-\). Here, \(B^-\) is the integrity component of \(B\) (similarly, \(B^+\) is the confidentiality component of \(B\)). This label means that Alice fully trusts host alice (with both confidentiality and integrity), while Bob trusts host alice to execute the program correctly, but does not trust the host with his secret data.

\begin{verbatim}
1 host alice: (A \land B^-)
2 host bob : (B \land A^-)
3
4 val a1, a2, a3 = input int alice
5 val b1, b2, b3 = input int bob
6 val a = min(a1, a2, a3)
7 val b = min(b1, b2, b3)
8 val b_richer = declassify a < b to (A \cap B)
9 output b_richer to alice, bob
\end{verbatim}

\hspace{1cm}
Figure 3. Guessing game, where Alice attempts to guess Bob’s secret number. Viaduct uses zero-knowledge proofs so Alice learns nothing more than whether her guesses are correct. Most labels in this code can be inferred automatically.

All variables and expressions in Viaduct carry a security label, which is derived from the possible flows of information in the program. The variables in lines 4–7 carry the same label as their respective hosts, since they only involve data local to that host. However, the comparison \( a < b \) involves both hosts’ private data, so has the higher security label \( A \land B \). This label corresponds to data that is secret to and trusted by both principals. Since \( A \land B \) corresponds to secret data, we require an explicit declassification to the label \( A \land B \), which describes data that both hosts can see and trust.

During protocol selection (§4), Viaduct chooses cryptographic protocols to securely and efficiently execute our example. The central idea that allows Viaduct to select protocols automatically is that the security guarantees of protocols can also be captured by labels. Neither Alice nor Bob alone has enough authority to be responsible for the comparison, so Viaduct generates the following distributed implementation: Alice and Bob compute their respective minima locally but perform the comparison \( a < b \) in semi-honest MPC. A semi-honest MPC protocol works here because the authority labels assigned to the hosts indicate that Alice and Bob trust each other’s hosts for integrity. Without that assumption, Viaduct is instead forced to select another protocol such as maliciously secure MPC.

There are typically multiple ways to assign protocols to a given program expression. For example, the computation of Alice’s minimum on line 6 could be securely performed in MPC, but since the computation requires the authority of Alice alone, it is cheaper yet still secure to do the computation locally on Alice’s machine. Using its cost estimator, Viaduct compiles the optimal program described above.

After protocol selection, Viaduct outputs a distributed program which captures the required cryptography to execute the source program. Hosts can execute this distributed program using Viaduct’s runtime system.

**Guessing Game.** Figure 3 presents a contrasting example. Here, Alice and Bob have security labels \( A \) and \( B \) respectively, modeling a malicious corruption scenario. Since they do not trust each other to execute the program correctly, semi-honest MPC is not applicable. Bob inputs a number \( n \), and Alice has five attempts to guess the number. Since Bob’s input initially has label \( B \), it must first be endorsed to the label \( B \land A^\top \), raising integrity so that Bob cannot unilaterally modify the value. This endorsement requires a cryptographic mechanism to protect the integrity and secrecy of variable \( n \) throughout program execution.

Viaduct synthesizes a program in which Bob commits to \( n \) so that its value remains secret to Alice but Bob cannot later lie about the committed value. The statement \( n == tguess \) is computed by having Bob send a zero-knowledge proof (ZKP) to Alice, so that Alice can trust the outcome but learns no additional information. All other variables are replicated in plaintext across the two hosts.

These examples show that Viaduct is general, as it treats protocols such as MPC and ZKP uniformly.

### 2.1 Specifying Security Policies

In Viaduct, security policies capture a notion of authority. Policies are represented by *principals*, formulas composed of conjunctions and disjunctions over a set of base principals \( \{A, B, C, \ldots \} \) and two special principals \( \emptyset \) and \( 1 \). Principal \( \emptyset \) represents maximal authority and corresponds to the conjunction of all base principals; principal \( 1 \) represents minimal authority and corresponds to the disjunction of all base principals. We distinguish authority over *confidentiality* and over *integrity*. The security requirements of information are thus characterized by *labels* consisting of pairs \( \langle p_1, p_2 \rangle \) of two principals \( p_1 \) and \( p_2 \), for confidentiality and integrity respectively.

A conjunction of principals \( p_1 \land p_2 \) represents combined authority. For confidentiality, this means the principal is allowed to read data that \( p_1 \) may read and also data that \( p_2 \) may read. For integrity, the conjunction may influence data that \( p_1 \) may influence, and also data that \( p_2 \) may influence. A disjunction \( p_1 \lor p_2 \) corresponds to common authority, which may read or influence exactly the data that either \( p_1 \) and \( p_2 \) may individually.

Principals carry a natural partial order based on their authority. We write \( p_1 \Rightarrow p_2 \) to mean \( p_1 \) “acts for”, or is at least as trusted as, \( p_2 \). This relation coincides with logical implication: for example, \( p_1 \land p_2 \Rightarrow p_1 \) and \( p_1 \Rightarrow p_1 \lor p_2 \).

It is convenient to have syntax that works over both components of labels simultaneously. So, we extend \( \emptyset, 1, \land, \lor, \Rightarrow \),
These projections are defined formally as follows:

\[ (p_c, p_i) \rightarrow \equiv (p_c, 1) \quad (p_c, p_i) \leftarrow \equiv (1, p_i). \]

The reflection operator [48] swaps the two components:

\[ \overline{\exists}(p_c, p_i) \equiv (p_c, p_i) \]

Viaduct programs assign labels to hosts to indicate the amount of trust placed in them, but there are also labels on data. The important insight, borrowed from FLAM [4], is that the same set of labels can be used to talk about both authority and information flow. When placed on data, a label takes on an information flow interpretation, specifying the minimum authority required to read and influence that data. As in FLAM, standard operations from information flow literature can be reformulated in terms of authority:

\[ \ell_1 \sqsubseteq \ell_2 \iff \ell_2^\rightarrow \Rightarrow \ell_1^\rightarrow \quad \text{and} \quad \ell_1^{-\rightarrow} \Rightarrow \ell_2^{-\rightarrow} \quad \text{(flows to)} \]

\[ \ell_1 \sqcup \ell_2 \equiv (\ell_1 \land \ell_2)^- \land (\ell_1 \lor \ell_2)^- \quad \text{(join)} \]

\[ \ell_1 \sqcap \ell_2 \equiv (\ell_1 \lor \ell_2)^- \land (\ell_1 \land \ell_2)^- \quad \text{(meet)} \]

The flows-to relation \( \ell_1 \sqsubseteq \ell_2 \) orders information flow policies: it means label \( \ell_1 \) is more permissive about the use of information than \( \ell_2 \). The join \( \ell_1 \lor \ell_2 \) is more restrictive than both \( \ell_1 \) and \( \ell_2 \), and the meet \( \ell_1 \sqcap \ell_2 \) is more permissive than either \( \ell_1 \) or \( \ell_2 \). The most restrictive label—that of completely secret, untrusted data—is \( 0^\rightarrow \equiv (0, 1) \), and the least restrictive (public, trusted data) is \( 0^- \equiv (1, 0) \).

2.2 Threat Model

Compiled programs run in a distributed setting in which each host executes a single thread concurrently with other hosts. Hosts communicate via message passing over secure, private, asynchronous channels. There is no shared memory that spans multiple hosts. We assume the attacker cannot observe wall-clock timing. Additionally, we are not concerned with availability, so the attacker can halt execution at any time.

In the setting of Viaduct, there is no single notion of an attacker. For example, in the historical millionaires problem, neither Alice nor Bob fully trust the other. To Alice, Bob is a potential attacker; Alice expects her security requirements to be met as long as the behavior of Bob’s (partially trusted) host is accurately described by the label assigned to it \( (B \land A^-) \). Conversely, to Bob, Alice is a potential attacker. Hence, we are concerned with security versus all possible attackers.

We model the power of an attacker using a label. The attacker can read the data on a host if the confidentiality of the attacker label is at least as trusted as that of the host, and can change data and code on the host if the integrity of the attacker label is at least as trusted as that of the host. We do not consider unreasonable attack scenarios in which a host has compromised integrity but still enforces confidentiality.\(^2\)

For example, in the historical millionaires’ problem, there are five interesting corruption scenarios: no corrupted hosts; Alice has corrupted confidentiality; Bob has corrupted confidentiality; both have corrupted confidentiality; or both Alice and Bob are fully corrupted. The full corruption of a single host is not possible because the hosts trust each other, so if the integrity of one is corrupted then the other’s integrity must be corrupted also.

2.3 Label Inference

Viaduct selects a protocol for every piece of data and computation in the program based on their authority requirements, represented as labels. Intuitively, program components must be executed by protocols with enough authority to defend the confidentiality of host inputs and the integrity of host outputs. These authority requirements are captured formally by a type system (§3.1), and Viaduct uses a novel inference algorithm (§3.2) to compute for all program components the minimum-authority labels that still respect the information flow constraints on the program.

The only required label annotations on Viaduct programs are the authority labels on host declarations and labels on declassify/endorse expressions—all labels on variables can be elided, making annotation burden low. As we show in our evaluation, in practice these required annotations are enough to capture programmer intent: minimally annotated programs compile to the same distributed programs as their fully annotated versions.

2.4 Protocol Selection

After label inference, Viaduct performs protocol selection, which assigns a protocol to compute and store each subexpression and variable. Protocols encompass storage and computation performed “in the clear” as well as cryptographic mechanisms such as commitments, MPC and zero-knowledge proofs.

Each protocol \( P \) carries an associated authority label \( L(P) \), which approximates the security guarantees the protocol provides. Given a program component with minimum authority requirement \( \ell \), protocol selection only assigns \( P \) to execute that component if \( L(P) \Rightarrow \ell \) —that is, if \( P \) meets the authority requirement for the program component.

Intuitively, given a program \( s \) and protocol \( P \), we may imagine an ideal functionality \( P^\star \) (in the style of UC [9]) which executes the program fragments of \( s \) that are assigned to \( P \). The fragments of \( s \) that are assigned to \( P \) may depend on the computational abilities of \( P \). For example, if \( P \) is a

\(^2\)The semi-honest and malicious threat models common in cryptography correspond to corrupting only hosts’ confidentiality and corrupting both hosts’ confidentiality and integrity respectively.
commitment protocol, then \( P^s \) is only able to store values but not perform any computations. If \( P \) is an MPC protocol, then \( P^s \) can execute computations that can be translated into circuits—the standard interface for MPC implementations.

\( P^s \) guarantees that the storage and computation it performs are protected at label \( \mathbb{L}(P) \). In particular, the adversary cannot observe storage or computation performed by \( P^s \) unless its confidentiality is at least \( \mathbb{L}(P) \); dually, the adversary cannot influence storage or computation performed by \( P^s \) unless its integrity is at least \( \mathbb{L}(P) \).

Examples of protocols and their corresponding security labels are given in Figure 4. Following the above intuition for the security of functionalities \( P^s \), the authority label of protocols are determined to be the least authority required of the adversary to corrupt the protocol (in confidentiality or integrity). We explain the example protocols below:

- **Local**. No cryptography is performed, and data is stored and computations performed on host \( h \) in the clear. It provides exactly the authority of \( h \).

- **Replicated**. Data and computations are replicated on all hosts in set \( H \), and replicated data is checked for equality when necessary. This protocol provides confidentiality \( \mathbb{V}_{h \in H} \mathbb{L}(h)^- \) since all hosts hold the plaintext value. It provides integrity \( \mathbb{W}_{h \in H} \mathbb{L}(h)^- \) since all hosts must corrupt their local values for the value to be globally corrupted. Together, these labels form the label \( \bigcap_{h \in H} \mathbb{L}(h) \).

- **Commitment**. Data is stored on \( h_p \) and commitments are placed on \( h_v \). Commitments are computationally inexpensive but usually no computations can be performed with them. Commitments increase integrity without sacrificing confidentiality. Its confidentiality is \( \mathbb{L}(h_p)^- \) since only \( h_p \) holds the plaintext value, while \( h_v \) only holds a commitment. Its integrity is \( (\mathbb{L}(h_p) \land \mathbb{L}(h_v))^- \) for the same reason as for replication.

- **ZKP**. A zero-knowledge proof protocol where \( h_p \) is the prover and \( h_v \) is the verifier. The prover computes over its private data and sends the result to the verifier, along with a proof that attests the value computed is correct. The proof reveals nothing about the private data except what can be gleaned from the result itself. Zero-knowledge proofs provide the same authority as commitments, for essentially the same reason: the prover holds all secret information and performs all computation, while the verifier only holds information which allows it to believe in the correctness of the result, but nothing more.

- **MAL-MPC**. A corrupt-majority, maliciously secure multiparty computation protocol performed by hosts \( H \). The protocol allows hosts to jointly perform a computation over their private inputs, keeping these inputs secret to the other hosts and revealing only the result. The label \( \bigcap_{h \in H} \mathbb{L}(h) \) reflects that the confidentiality (resp., integrity) of data computed in MPC is compromised only if all participating hosts have compromised confidentiality (resp. integrity).

- **SH-MPC**. A corrupt-majority, semi-honest secure multiparty computation protocol performed by hosts \( H \). While the combined authority label is complex, its confidentiality and integrity projections are easy to understand. The integrity is equal to \( \mathbb{V}_{h \in H} \mathbb{L}(h)^- \), since the integrity of the MPC computation may be compromised if any host behaves maliciously. The confidentiality is equal to

\[
\mathbb{V}_{h \in H} \mathbb{L}(h)^- \lor \bigcap_{h \in H} \mathbb{L}(h)^-.
\]

The first disjunct captures the fact that confidentiality guarantees are discarded if the integrity of any host is compromised. The second disjunct states that, if all hosts follow the protocol correctly, the adversary can only learn the state of intermediate MPC computations if all hosts have corrupted confidentiality. Overall, this means that in order to compromise confidentiality guarantees of semi-honest MPC, either the integrity of any host or the confidentiality of all hosts must be compromised.

In particular, for the historical millionaires’ example, the labels of SH-MPC(\( alice, bob \)) is \( A \land B \). This is because hosts \( alice \) and \( bob \) are both assumed to have the high integrity of \( (A \land B)^- \). If \( alice \) and \( bob \) only have their own integrity, however, then the label is computed to be \( A \lor B \). The protocol only has enough authority to perform computations over data public to both hosts, and neither host trusts the result. Indeed, semi-honest MPC offers little to no benefit if any host has lower integrity than any other.

### 2.5 Runtime

Viaduct provides a modular runtime system for executing compiled distributed programs, implemented as an interpreter. All hosts run the interpreter with the same compiled program, which then executes each host’s portion of the program. During execution, the interpreter calls out to back ends implementing the cryptographic mechanisms used in the program. Back ends translate computations in the source language into their cryptographic realizations. For instance,
the back ends for MPC and ZKP in our implementation build a circuit representation of the program as it executes.

Protocol back ends can send data to and receive data from each other, supporting the composition of protocols. Source-level declassification and endorsement induce this communication. For example, in Figure 2 on line 8, the computation $a < b$ is declassified from label $A \wedge B$ to $A \cap B$. This declassification causes the MPC protocol between Alice and Bob to execute its stored circuit for this comparison, and to output the result in cleartext.

Figure 5 shows the execution of the program compiled by Viaduct for the historical millionaires’ problem. The program runs as follows. (1) First, the cleartext back ends on Alice and Bob’s machines receive input locally and compute their respective minima. The back ends send the minima as secret inputs to their respective MPC back ends, which create input gates for these inputs. (2) Next, the MPC back ends on Alice and Bob’s machines each create an operation gate that compares Alice and Bob’s secret inputs. The back ends jointly execute the circuit with the comparison result as output, which they send to their respective cleartext back ends. (3) Finally, the cleartext back ends on Alice and Bob’s machines both receive from their MPC back ends and output the result.

### 3 Source Language

The syntax for Viaduct’s source language, a simplified version of the surface language, is given in Figure 6. The language supports base types such as booleans and integers, along with their usual operators. Surface-level assignables (val and var declarations) and arrays are uniformly represented as data types, a restricted form of objects. Like regular objects, they are created using constructors (new declarations) and contain methods. For simplicity, we only include three data types: immutable/mutable cells, which model surface-level assignables, and arrays. Arrays are dynamically sized but statically allocated: the size of an array can depend on values known only at run time, but array references cannot be rebound to different names or stored in arrays.

We distinguish between fully evaluated atomic expressions $a$, and expressions $e$ that evaluate to values and may have side effects. Methods include get and set operations for both mutable cells and arrays (for which they take an index as an extra argument). Input/output expressions allow programs to interact with hosts. The declassify expression marks locations where private data is explicitly allowed to flow to public data, while the endorse expression marks locations where untrusted data is explicitly allowed to influence trusted data.

Statements consist of let-bindings, assignable declarations, as well as the usual conditionals, loops, and sequential composition. Temporaries bind values while assignables bind instances of data types. We require all intermediate computations to be let-bound by a temporary, enforcing a variant of A-normal form [18]. We use the more general loop-unti...
simplifying the conversion to A-normal form. A break statement (break $b$) includes an identifier $b$ that names the loop it breaks out of.

### 3.1 Label Checking

Viaduct’s type system enforces secure information flow in a standard way. The type system serves two purposes. First, it helps programmers ensure there are no unintended information flows: secrets are not leaked to and data is not corrupted by unauthorized principals. Second, it specifies what labels can be assigned to variables and expressions that the user did not explicitly annotate.

Figure 7 presents label checking rules for expressions and selected statements. Expressions are checked by the judgment $\Gamma; pc \vdash e : \ell$, which means that $e$ has label $\ell$ under the context on the left. Here, $\Gamma$ is a finite partial map from temporaries, assignables, or loop names to labels:

$$\text{Label Contexts} \quad \Gamma ::= \cdot | \Gamma, t : \ell \mid \Gamma, x : \ell \mid \Gamma, b : \ell$$

The program counter label $pc$ is a standard way to prevent implicit flows of information via control flow [43]. The rules for method calls and input/output expressions differ from those in standard security-typed languages in that they also include premises with $pc$ checks. These checks are required because these expressions may induce communication between hosts, and hosts may learn secrets based on which requests they receive. Prior work that targets the distributed setting contains similar checks to control read channels [51].

Statement checking rules have the form $\Gamma; pc \vdash s$; they are largely standard [43]. Because we assume attackers cannot observe timing nor analyze traffic, the rule for conditional statements does not require branches to have the same timing behavior or effects (e.g., method calls, input/output).

### Nonmalleable Information Flow Control

Information flow type systems typically aim to enforce a compositional security property such as noninterference [23]. Noninterference is a strong property but it is too restrictive for practical applications, which usually have a more nuanced policy for secure information flow. Hence, like most languages supporting information flow control (e.g., [6, 36, 39]), Viaduct allows programmers to signify the exceptions to a noninterference policy through downgrading expressions.

Downgrading enables information flows that would violate noninterference, so it can be dangerous. This is especially true in the distributed setting, where storage and computation can be performed by hosts that one does not fully trust. Downgrading confidentiality (declassification) allows secret information to be treated as public information—a necessity for many applications, but doing so might allow a corrupted host to control when information is released or what information is released. Downgrading integrity (endorsement) allows untrusted information to be treated as trusted information, but might enable a corrupted host to trick an honest one into accepting mauld secrets.

The property of nonmalleable information flow control (NMIFC) [11] prevents both of these abuses of downgrading by combining two properties: robust declassification [50] and transparent endorsement [11]. Robust declassification requires that principals to which data is declassified could not have influenced either the decision to declassify or the data itself. Meanwhile, transparent endorsement prevents trusting mauld secrets by ensuring that information can only be endorsed if the providing principal can read it.

The declassification and endorsement rules in Figure 7 enforce NMIFC using the reflection operator $\mathcal{X}$ (§2.1). The rules prevent the program from downgrading information with compromised labels [48], in which confidentiality exceeds integrity. These rules generate authority requirements that prevent the Viaduct compiler from placing data and computation on insufficiently trustworthy hosts. For example, consider a program where a server releases secret information to a client when the client guesses the correct password:

```plaintext
host server: (5), client: (1)
val info: int{5}, pw: int{5}, guess: int{1}
if (declassify (pw == guess) to (1))
  output (declassify info to (1)) to client
```

This program violates robust declassification, because the decision to declassify $info$ depends on (low-integrity) $guess$. Without the restrictions on downgrading, Viaduct could compile the program to store the guard $pw == guess$ (with label 1) on the client. The client could simply claim to the server that its guess is correct! For this program to type-check with NMIFC, endorsement is needed to make the guard high-integrity. A naive programmer might think to endorse the entire guard, but this (nontransparent) endorsement could still be compiled in a way that lets an untrusted host supply its value. The correct solution is to explicitly endorse $guess$ before declassifying the comparison; since $guess$ is not secret, the endorsement is transparent. The resulting labels correctly force Viaduct to put the comparison on the server.

### 3.2 Label Inference

Checking secure information flow is not enough; for protocol selection, the compiler also needs the labels of all expressions. We present an algorithm to infer these labels. As in prior work on inferring information flow labels [36, 39], information flow checking reduces to a system of flows-to ($\Sigma$) constraints over label constants and label variables. Type inference collects these premises from Figure 7, and generates fresh label variables for labels that appear in a premise of a rule but not its conclusion (e.g., $pc'$ in the rule for if statements). The inference algorithm finds a label-variable assignment that satisfies all the constraints, if possible.
The algorithm computes the minimum-authority solution, the choice of labels requiring the least amount of confidentiality and integrity for each component. Minimum-authority labels are desirable because higher authority is achieved only through more trust or costly cryptography.

First, we translate the flows-to (⊇) constraints over labels, which appear in rule premises, to acts-for (⇒) constraints over the underlying label components as shown in Figure 8. Here, C(ℓ) and I(ℓ) are functions that project the confidentiality and integrity components, respectively, of label ℓ. These components are constants p when the label is known, and variables L otherwise.

We then adapt the algorithm of Rehof and Mogensen [41] for iteratively solving semilattice constraints. All principal variables are initialized to 1 and unsatisfied constraints are used to update variables repeatedly, until a fixed point is reached, according to the rules in Figure 9. Constraints of the form \( L_1 \Rightarrow L_2 \) or \( L_1 \Rightarrow L_2 \lor L_3 \) are used to perform the corresponding update.

However, the rules in Figure 8 can also generate constraints of the form \( L_1 \land p_2 \Rightarrow L_3 \), arising from the typing rule for robust declassification. The term \( p_2 \) is always a constant since Viaduct requires annotations on declassify operations, so the value of \( L_1 \) can be updated safely to \( p_2 \rightarrow L_3 \), which denotes the weakest authority \( p \) such that \( p \land p_2 \Rightarrow L_3 \). When a lattice supports the \( \rightarrow \) operation, it is a Heyting algebra [42], allowing each update rule to lower the left-hand-side variable to the minimum authority satisfying the constraint. Any free distributive lattice, such as our lattice of principals, is a Heyting algebra. We prove this fact, as well as the fact that iterative analysis always terminates with the minimum-authority solution, in the supplementary technical report [2].
Viaduct

Figure 10. Selected rules for the validity of a protocol assignment.

\[
\begin{align*}
\Pi \models e : P & \quad \Pi \models t : P \\
\Pi \models a_i : \Pi(x) & \quad \Pi \models x.m(a_1, \ldots, a_n) : \Pi(x) \\
\Pi \models \text{input}_h a : \text{Local}(h) & \quad \Pi \models \text{output} a \rightarrow h : \text{Local}(h) \\
\end{align*}
\]

\[
\begin{align*}
\Pi \models t = e \text{ in } s & \quad \Pi \models t = e \text{ in } s \\
\Pi \models \text{new } x = D(a_1, \ldots, a_n) \text{ in } s & \quad \Pi \models \text{new } x = D(a_1, \ldots, a_n) \text{ in } s \\
\Pi \models \text{if } a \text{ then } s_1 \text{ else } s_2 & \quad \Pi \models \text{if } a \text{ then } s_1 \text{ else } s_2 \\
\end{align*}
\]

Figure 11. Protocols and hosts involved in the execution of a statement. Here, \(\text{hosts}(P)\) is the set of hosts that protocol \(P\) runs on, which is specified individually for each protocol.

\[
\Pi(s) : 2^P \quad \text{hosts}(\Pi(s)) : 2^P
\]

\[
\begin{align*}
\Pi(\text{let } t = e \text{ in } s) & = \Pi(t) \cup \Pi(s) \\
\Pi(\text{new } x = D(a_1, \ldots, a_n) \text{ in } s) & = \Pi(x) \cup \Pi(s) \\
\Pi(\text{if } a \text{ then } s_1 \text{ else } s_2) & = \Pi(s_1) \cup \Pi(s_2) \\
\Pi(b : \text{loop } s) & = \Pi(s) \\
\Pi(\text{break } b) & = \Pi(b : \text{loop } s)
\end{align*}
\]

\[
\text{hosts}(\Pi(s)) = \bigcup_{P \in \Pi(s)} \text{hosts}(P)
\]

Figure 12. Abstract cost model.

\[
\text{cost}(\Pi, \text{let } t = e \text{ in } s) = \\
c_{\text{exec}}(\Pi(t), e) + \sum_{P \in \text{readers}(\Pi, t,s)} c_{\text{comm}}(\Pi(t), P) + \text{cost}(\Pi, s)
\]

\[
\text{cost}(\Pi, \text{if } a \text{ then } s_1 \text{ else } s_2) = \max(\text{cost}(\Pi, s_1), \text{cost}(\Pi, s_2))
\]

\[
\text{cost}(\Pi, b : \text{loop } s) = W_{\text{loop}} \times \text{cost}(\Pi, s)
\]

\[
\text{cost}(\Pi, s_1; s_2) = \text{cost}(\Pi, s_1) + \text{cost}(\Pi, s_2)
\]

\[
\text{cost}(\Pi, s) = 0 \text{ otherwise}
\]

4.2 Cost of Protocol Assignments

There can be many valid protocol assignments that securely realize a source program. To select an optimal assignment, Viaduct attributes a cost to each assignment using an abstract cost model, shown in Figure 12. Developers can instantiate the abstract model by modifying the customizable cost estimator, which specifies \(c_{\text{exec}}(P, s)\), the cost of executing statement \(s\) in protocol \(P\); \(c_{\text{comm}}(P_1, P_2)\), the cost of communicating between \(P_1\) and \(P_2\); and the global constant \(W_{\text{loop}}\), the number of times a loop is assumed to execute when its iteration count is not statically known.

Our implementation configures \(c_{\text{exec}}\) to assign a small cost to executing "in the clear" and a large cost to the use of cryptography, so the compiler avoids the use of cryptography except when required for security. We also configure the communication cost \(c_{\text{comm}}\) to minimize data movement. For example, a frequently accessed public variable would be replicated on two hosts so that each host has a local copy. Placing the variable only on one of the hosts could reduce storage cost but entails frequently sending its value to the other host.
4.3 Computing an Optimal Protocol Assignment

To compute an optimal protocol assignment given a program $s$, the Viaduct compiler constructs a constrained optimization problem over the following sets of variables:

- **Assignment variables ($\alpha_i$)**. These represent the protocols that execute let-bindings or declarations.
- **Cost variables ($\beta_j$)**. These represent the cost of executing let-bindings or declarations.
- **Participating host variables ($y_{1,j}$)**. These are true if host $j$ is participating in the execution of a statement $i$.

The compiler generates a set of constraints \{\phi_1, \ldots, \phi_n\} over these assignment, cost, and participating host variables, as well as an expression $\beta_i$ capturing the cost of $s$ as in Figure 12. These constraints are drawn from a grammar consisting of logical connectives, an equality predicate between assignment variables and protocols, and an equality predicate between cost variables and cost expressions. The compiler uses an off-the-shelf solver to find a solution for assignment variables $\alpha_i$ and participating host variables $y_{1,j}$ such that all constraints $\{\phi_1, \ldots, \phi_n\}$ are satisfied and $\beta_i$ is minimized.

Given the set of valid protocol assignments $VA$ for $s$ such that $VA = \{\Pi \mid \Pi \models s\}$, this solution for the assignment variables corresponds to a protocol assignment $\Pi_{opt}$ such that

$$\Pi_{opt} = \arg \min_{\Pi \in VA} \text{cost}(\Pi, s).$$

**Protocol Factory.** To construct the optimization problem, the compiler draws the set of available protocols from the customizable protocol factory. Developers wishing to add new protocols to Viaduct must extend the protocol factory so that the compiler can generate assignments with these protocols during protocol selection.

The protocol factory defines a function $\text{viable} : T \cup T \rightarrow 2^P$ that returns a set of viable protocols that can execute a let-binding or declaration. This allows developers to specify limitations regarding the use of particular protocols. For example, commitment protocols may be unable to compute over commitments. Other protocols may lack support for certain operators.

**Example.** Consider the following source program to be executed by hosts $a$ and $b$:

```
let $t_1 = 1 + 1 \text{ in } t_2 = t_1 \times 2 \text{ in skip}
```

and the following data from the compiler’s label inference phase and extension points:

1. $\text{viable}(t_1) = \{P_1, P_2, P_4\}$, $\text{viable}(t_2) = \{P_1, P_2\}$
2. $L(P_1) \Rightarrow L(t_1)$, $L(P_2) \Rightarrow L(t_1)$, $L(P_4) \not\Rightarrow L(t_1)$
3. $L(P_1) \Rightarrow L(t_2)$, $L(P_2) \Rightarrow L(t_2)$
4. $\text{hosts}(P_1) = \{a\}$, $\text{hosts}(P_2) = \{b\}$, $\text{hosts}(P_4) = \{a, b\}$
5. $c_{\text{exec}}(P_1, -) = 5$, $c_{\text{exec}}(P_2, -) = 5$, $c_{\text{exec}}(P_3, -) = 3$
6. $c_{\text{comm}}(P_1, P_1) = 0$, $c_{\text{comm}}(P_2, P_2) = 2$
7. $\text{comm}(P_1, P_1) \lor \text{comm}(P_2, P_2)$
8. $\text{comm}(P_3, P_2) \lor \text{comm}(P_1, P_2)$

Then the compiler constructs the problem of minimizing cost $\beta_1 + \beta_2$ while satisfying the following constraints:

$$\begin{align*}
(\alpha_1 &= P_1 \lor \alpha_1 = P_3) \land (\alpha_2 = P_1 \lor \alpha_2 = P_2) \\
\alpha_1 &= P_1 \rightarrow (y_{1,1} \land \neg y_{1,2} \land \beta_1 = 5) \\
\alpha_1 &= P_3 \rightarrow (y_{1,1} \land y_{1,2} \land \beta_1 = 3) \\
\alpha_2 &= P_1 \rightarrow (y_{2,1} \land \neg y_{2,2} \land \alpha_1 \neq P_3 \land (\alpha_1 = P_1 \rightarrow \beta_2 = 5 + 0)) \\
\alpha_2 &= P_2 \rightarrow (\neg y_{2,1} \land y_{2,2} \land \alpha_1 \neq P_1 \land (\alpha_1 = P_3 \rightarrow \beta_2 = 3 + 2))
\end{align*}$$

Note that $\alpha_1, \beta_1, y_{1,1}$, and $y_{1,2}$ are variables associated with $t_1$ while $\alpha_2, \beta_2$ and $y_{2,1}, y_{2,2}$ are variables associated with $t_2$. The first constraint bounds the possible values of assignment variables $\alpha_1$ and $\alpha_2$ and is generated from the viable protocols returned by the protocol factory. Viable protocols that do not meet authority requirements are filtered out, so $P_4$ is not a possible value for $\alpha_1$. The rest of the constraints describe the relationship between protocol assignments, participating hosts, possible protocol compositions, and cost. From this optimization problem the compiler then computes the optimal assignment $\Pi_{opt}$ where $\Pi_{opt}(t_1) = P_3$ and $\Pi_{opt}(t_2) = P_2$.

5 Viaduct Runtime

Once it has computed a protocol assignment, the Viaduct compiler outputs a program where every let-binding and assignable declaration is annotated with the protocol that will execute it. This annotated program can be executed by the Viaduct runtime, which consists of an extensible interpreter that interacts with a set of protocol back ends, each of which implement a set of protocols. The interface for protocol back ends is straightforward: back ends must implement methods to execute let-bindings and assignable declarations, and methods to communicate with other protocol back ends.

Each host runs a copy of the interpreter with the annotated program as input. For each statement, the interpreter checks whether the host participates in its execution, as defined by $\text{hosts}(\Pi, -)$—if not, the statement is treated like skip. If a host participates in executing a let-binding or a declaration, the interpreter calls the back end for the protocol assigned to the statement. To execute a conditional, the host retrieves the cleartext value of the guard from the protocol back end that stores it, and executes the appropriate branch. The validity rules for protocol assignments ensure the host is allowed to see the cleartext value, and that it is able to retrieve it.

5.1 Protocol Composition

The protocol back end executing a let-binding must send the computed value to back ends executing statements that read the bound temporary. How one back end sends a value to another depends on the protocols involved. For example, a statement executed in $\text{Replicated}(h_1, h_2)$ reading a

---

Note that participating host variables are unused here, but in general they encode the guard visibility constraint for conditionals.
### Table 1. Selected examples of protocol composition.

<table>
<thead>
<tr>
<th>Sending protocol (s)</th>
<th>Receiving protocol (r)</th>
<th>Communication</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local($h_1$)</td>
<td>SH-MPC($h_1$, $h_2$)</td>
<td>$(s, h_1) \rightarrow (r, h_1)$</td>
<td>create input gate for MPC circuit</td>
</tr>
<tr>
<td>Local($h_2$)</td>
<td>Commitment($h_p$, $h_v$)</td>
<td>$(s, h_2) \rightarrow (r, h_p)$</td>
<td>create commitment</td>
</tr>
<tr>
<td>Replicated($h_1$, $h_2$)</td>
<td>SH-MPC($h_1$, $h_2$)</td>
<td>$(s, h_1) \rightarrow (r, h_1), (s, h_2) \rightarrow (r, h_2)$</td>
<td>read replicated data</td>
</tr>
<tr>
<td>SH-MPC($h_1$, $h_2$)</td>
<td>Replicated($h_1$, $h_2$)</td>
<td>$(s, h_1) \rightarrow (r, h_1), (s, h_2) \rightarrow (r, h_2)$</td>
<td>execute circuit and reveal output</td>
</tr>
<tr>
<td>Commitment($h_p$, $h_v$)</td>
<td>Local($h_v$)</td>
<td>$(s, h_p) \rightarrow (r, h_v)$</td>
<td>open commitment</td>
</tr>
<tr>
<td>ZKP($h_p$, $h_v$)</td>
<td>Local($h_v$)</td>
<td>$(s, h_v) \rightarrow (r, h_v)$</td>
<td>send result and proof to verifier</td>
</tr>
</tbody>
</table>

Figure 13 shows a table of selected compositions and the messages that constitute them. The table illustrates our insight that protocol composition is a general abstraction to represent the use of cryptographic mechanisms. The creation of a commitment and its opening; the execution of an MPC circuit and the revealing of its output; a prover sending a zero-knowledge proof to a verifier—all of these are captured by a composition of one protocol with another.

### 6 Implementation

We implemented the Viaduct compiler in about 20 KLoC of Kotlin code, which includes code for the parser, the label constraint solver, protocol selection, and the runtime system. The code written against the compiler’s extension points—the protocol factory, the protocol composer, the cost estimator, and the protocol back ends—runs to about 4 KLoC. Viaduct uses the Z3 SMT solver [14] to solve the optimization problem generated during protocol selection.

The compiler supports the more liberal surface syntax seen in Figure 2 and Figure 3, as well as functions with bounded polymorphism on parameter labels. The compiler specializes functions at each call site, allowing different compiled implementations for the same function.

We implemented four protocol back ends for Viaduct:

**Local/Replicated.** The cleartext back end executes code in Local and Replicated protocols. It maintains a store for objects that directly represent the temporaries and assignables of the source program. Computations performed by the cleartext back end are executed directly.

**SH-MPC.** This back end links Viaduct to ABY, a library for two-party semi-honest MPC [16]. It maintains a store of gate objects that represent circuit components executed by ABY. Computations performed by the back end build gate objects that represent the operation performed (e.g., an addition in the source program creates an ADD gate).

The ABY framework supports execution of circuits in three different schemes—arithmetic sharing, boolean sharing, and...
Yao’s garbled circuits—as well as conversions between these, allowing for execution of mixed-protocol circuits. Viaduct represents each scheme as a separate protocol, but all three are implemented by a single back end. To generate efficient mixed circuits, we follow Demmler et al. [16] and Ishaq et al. [28] and estimate inputs to the cost estimator by measuring execution time of individual operations under a particular scheme and conversions between schemes. We perform measurements for two settings: low-latency, high-bandwidth (LAN), and high-latency, low-bandwidth (WAN). Thus the cost estimator has two modes, each of which optimizes compiled programs for a specific network environment.

**Commitment.** This back end manages commitments, implemented using SHA-256 hashes of data along with a nonce. The back end for the commitment creator maintains a store of cleartext values along with metadata for commitments. The back end for the commitment receiver maintains the set of commitments, as hashes. The commitment back end cannot support computation.

**ZKP.** This back end links to libsnark [1], a library for zkSNARKs (zero-knowledge Succinct Non-interactive Arguments of Knowledge). This back end maintains a store of circuit gate objects. The prover and verifier both manage cleartext values for the public inputs to the proof, while only the prover manages cleartext values for the secret inputs. To ensure the prover cannot modify secret inputs mid-execution, all secret inputs are “committed” by sending their hash to the verifier. All proofs that use a secret input then include a clause that equates the input to the pre-image of the hash held by the verifier.

The libsnark library requires proving and verifying keys to be generated for each unique circuit before the protocol is executed. The current prototype requires a “dummy” run of the compiled program to generate these keys.

### 7 Evaluation

To evaluate Viaduct, we address these research questions:

- **RQ1:** Is Viaduct expressive enough?
- **RQ2:** Is its compilation performance acceptable?
- **RQ3:** Does it generate efficient distributed programs?
- **RQ4:** How much does label inference reduce the annotation burden for programmers?
- **RQ5:** What is the overhead of the runtime system?

Experiments used Dell OptiPlex 7050 machines with an 8-core Intel Core i7 7th Gen CPU and 16 GB of RAM. Note that for experiments involving time measurements (RQ2, RQ3, RQ5), the numbers reported are over 5 trials and the relative standard error is at most 6% of the sample mean.

---

**RQ1 - Expressiveness.** Figure 14 shows the benchmarks used for the experiments and the cryptography synthesized by Viaduct for each benchmark. Several are from prior work, rewritten in the Viaduct source language. Host configurations are either semi-honest, as in Figure 2, where hosts A and B trust each other for integrity; mutually distrusting as in Figure 5; or are “hybrid” configurations where A and B trust each other but host C is trusted by neither.

Our benchmarks show that Viaduct can compile programs whose security demands a variety of cryptographic mechanisms. With hybrid configurations (interval, bet), Viaduct combines MPC and ZKP to implement different components of a single distributed program. Code for selected benchmarks can be found in the supplemental technical report [2].

**RQ2 - Scalability of Compilation.** The two main phases of the Viaduct compiler are label inference and protocol selection. Our benchmarks indicate that the overhead of label inference is negligible: at most several hundred milliseconds. As seen in Figure 14, the overhead for protocol selection is more significant, but still on the order of several seconds for most benchmarks. The longest running benchmark, k-means, performs most of its computations in MPC. In this case, it may be harder to converge to the optimal solution since the solver generates a large mixed circuit, choosing between the three MPC schemes supported by ABY.

**RQ3 - Cost of Compiled Programs.** To show that Viaduct can compile efficient distributed programs, we chose a subset of our benchmarks requiring the use of MPC and compared the execution of optimal programs generated by Viaduct—for each benchmark, one optimized for local area networks (LAN) and another for wide area networks (WAN)—with naive protocol assignments that perform all computation in MPC. The naive ABY assignments use either boolean sharing or Yao garbled circuits, since arithmetic sharing can only perform arithmetic operations. We measured executions in a 1 Gbps LAN and simulated WAN (100 Mbps bandwidth and 50 ms latency). We configured ABY to use 32-bit integers and set its security parameter to 128 bits.

Figure 15 summarizes our results. For some benchmarks (HHI score, hist. millionaires, median, two-round bidding), computation can be securely moved from MPC to cleartext protocols, making execution much more efficient. Even for benchmarks that require computations to be almost entirely in MPC (bio. match, k-means), Viaduct chooses efficient mixed circuits that perform much better than the naive assignments entirely in boolean sharing or Yao circuits. Viaduct replicates the result in Büscher et al. [7] (which specifically targets the ABY framework) in choosing a mix of arithmetic and Yao circuits as optimal assignments for the two benchmarks from that paper, with the exception of the k-means benchmark in the WAN setting.
### Protocols Selection

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Description</th>
<th>Protocols</th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>battleship</td>
<td>model of the board game</td>
<td>RZ / RZ</td>
<td>Ann   12</td>
</tr>
<tr>
<td>bet</td>
<td>C bets who wins hist. millionaires b/w A &amp; B</td>
<td>CLRY / CLRY</td>
<td>Ann  7</td>
</tr>
<tr>
<td>biometric match</td>
<td>min distance b/w sample &amp; database (from [7])</td>
<td>ALRY / ALRY</td>
<td>Ann  8</td>
</tr>
<tr>
<td>guessing game</td>
<td>same as in fig. 3</td>
<td>RZ / RZ</td>
<td>Ann   16</td>
</tr>
<tr>
<td>HHI score</td>
<td>compute market concentration index (from [45])</td>
<td>ALRY / LRY</td>
<td>Ann  22</td>
</tr>
<tr>
<td>historical millionaires</td>
<td>same as in fig. 2 but with arrays</td>
<td>LRY / LRY</td>
<td>Ann  17</td>
</tr>
<tr>
<td>interval</td>
<td>A &amp; B compute interval of combined points, C attests point is in interval</td>
<td>RYZ / RYZ</td>
<td>Ann  45</td>
</tr>
<tr>
<td>k-means</td>
<td>cluster secret points from A &amp; B (from [7])</td>
<td>ARY / RY</td>
<td>Ann  82</td>
</tr>
<tr>
<td>k-means (unrolled)</td>
<td>k-means w/ 3 unrolled iterations</td>
<td>ARY / RY</td>
<td>Ann 174</td>
</tr>
<tr>
<td>median</td>
<td>compute median of A &amp; B’s lists (from [30])</td>
<td>RY / RY</td>
<td>Ann  36</td>
</tr>
<tr>
<td>rock-paper-scissors</td>
<td>A &amp; B commit to moves then reveal</td>
<td>CR / CR</td>
<td>Ann  56</td>
</tr>
<tr>
<td>two-round bidding</td>
<td>A &amp; B bid for a list of items</td>
<td>LRY / LRY</td>
<td>Ann  34</td>
</tr>
</tbody>
</table>

Figure 14. Benchmark programs. **Protocols** give the protocols used in the compiled program for either the LAN or WAN setting. Legend for protocols used: A, B, Y = ABY arithmetic/boolean/Yao sharing; C = Commitment; L = Local; R = Replicated; Z = ZKP. **Ann** gives the minimum number of label annotations needed to write the program. **Selection** gives the number of symbolic variables and run time in seconds for protocol selection, averaged across five runs.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>LAN</th>
<th>WAN</th>
<th>Comm</th>
<th>LAN</th>
<th>WAN</th>
<th>Comm</th>
<th>LAN</th>
<th>WAN</th>
<th>Comm</th>
</tr>
</thead>
<tbody>
<tr>
<td>bio. match</td>
<td>3.6</td>
<td>95.9</td>
<td>56.0</td>
<td>2.8</td>
<td>7.1</td>
<td>52.3</td>
<td>1.0</td>
<td>2.2</td>
<td>3.9</td>
</tr>
<tr>
<td>HHI score</td>
<td>0.8</td>
<td>9.7</td>
<td>7.0</td>
<td>0.5</td>
<td>1.6</td>
<td>2.7</td>
<td>0.3</td>
<td>1.1</td>
<td>0.5</td>
</tr>
<tr>
<td>hist. million.</td>
<td>1.0</td>
<td>90.6</td>
<td>4.8</td>
<td>0.6</td>
<td>1.6</td>
<td>3.1</td>
<td>0.3</td>
<td>0.7</td>
<td>0.005</td>
</tr>
<tr>
<td>k-means</td>
<td>56.5</td>
<td>696.1</td>
<td>1273.1</td>
<td>44.4</td>
<td>117.4</td>
<td>1051.3</td>
<td>17.7</td>
<td>35.8</td>
<td>180.0</td>
</tr>
<tr>
<td>median</td>
<td>11.5</td>
<td>1098.7</td>
<td>197.1</td>
<td>12.8</td>
<td>35.4</td>
<td>327.8</td>
<td>0.7</td>
<td>31.7</td>
<td>1.0</td>
</tr>
<tr>
<td>2-R bidding</td>
<td>17.3</td>
<td>184.7</td>
<td>233.0</td>
<td>17.8</td>
<td>184.5</td>
<td>233.0</td>
<td>3.1</td>
<td>155.5</td>
<td>4.7</td>
</tr>
</tbody>
</table>

Figure 15. Run time (in seconds) and communication (in MB) of select benchmark programs, averaged across five runs. **Bool** and **Yao** are naive assignments using boolean sharing and Yao sharing respectively to execute MPC computations. **Opt-LAN** and **Opt-WAN** are optimal assignments generated by Viaduct for the LAN and WAN setting respectively. Optimal time and communication for a benchmark and execution setting pair are in **bold**.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>LAN Time</th>
<th>LAN Slowdown</th>
<th>WAN Time</th>
<th>WAN Slowdown</th>
</tr>
</thead>
<tbody>
<tr>
<td>bio. match</td>
<td>0.4</td>
<td>150%</td>
<td>1.5</td>
<td>50%</td>
</tr>
<tr>
<td>HHI score</td>
<td>0.3</td>
<td>0%</td>
<td>1.0</td>
<td>10%</td>
</tr>
<tr>
<td>hist. million.</td>
<td>0.3</td>
<td>0%</td>
<td>0.7</td>
<td>0%</td>
</tr>
<tr>
<td>k-means</td>
<td>1.2</td>
<td>1380%</td>
<td>4.1</td>
<td>770%</td>
</tr>
<tr>
<td>median</td>
<td>0.5</td>
<td>40%</td>
<td>31.5</td>
<td>0%</td>
</tr>
<tr>
<td>2-R bidding</td>
<td>1.6</td>
<td>90%</td>
<td>154.7</td>
<td>0%</td>
</tr>
</tbody>
</table>

Figure 16. Run time (in seconds) of LAN-optimized benchmarks hand-written to use ABY directly and the slowdown of running the same benchmarks through the Viaduct runtime in LAN and WAN settings.

**RQ4 - Annotation Burden of Security Labels.** Security-typed languages add some annotation burden when writing programs. In practice, labels on host declarations and downgrading operations suffice to specify intended security policies in Viaduct programs. To substantiate this claim, we created two versions of each benchmark program. In one, every variable has a label annotation; in the other, “erased” version, all such labels are omitted.

For all benchmarks, Viaduct generates the same compiled program for the fully labeled and the erased versions. Although the inferred labels for the erased programs are not exactly the same as in their manually labeled counterparts, the differences do not affect the protocols chosen. The **Ann** column in Figure 14 counts label annotations on erased programs. This is the minimum number of annotations needed to write the program: effectively, the number of downgrades.
plus the number of host declarations, each of which need an authority label. The table shows that the annotation burden is low: most benchmarks need only a few label annotations.

**RQ5 - Overhead of Runtime System.** The Viaduct runtime introduces some overhead compared to using cryptographic libraries like ABY directly. To measure this overhead, we translated Viaduct’s LAN-optimized outputs for the MPC benchmarks in Figure 15 to directly use the ABY framework’s API. We then measured the performance of these hand-written programs in the LAN and WAN settings.

Figure 16 gives running times for the hand-written programs and the overhead of using the Viaduct runtime. For most benchmarks, the Viaduct runtime incurs an overhead of at most 150% in the LAN setting; the overhead is reduced to at most 50% in the WAN setting where network delay is a more significant factor. This overhead is due to the cost of interpretation and dynamic circuit generation, and can be eliminated by moving circuit generation to compile time.

The markedly larger overhead of the k-means benchmark is due to Viaduct recomputing intermediate results. The benchmark has 8 outputs; while Viaduct evaluates 8 smaller MPC circuits each with one output, the hand-written version evaluates one larger circuit with 8 outputs, taking advantage of shared intermediate computations. The compiler could, with additional analysis, determine when output gates can be grouped and executed in the same circuit. We leave this to future work.

8 Related Work

**Compilation to Cryptographic Protocols.** The idea of compiling a high-level program to a cryptographic protocol has been explored in the context of multiparty computation [27] (e.g., Fairplay [34], SCVM [32], ObliVM [33], OblivC [49], Wysteria [40], HYCC [7], SCALE-MAMBA [3]), and that of zero-knowledge proofs (e.g., Pinocchio [37], Gempetto [13], Buffet [46], jxsSNARK [31]). Earlier work is generally limited to the domain of a particular fixed cryptographic task (e.g., MPC or ZKP); Viaduct’s novelty is synthesizing efficient protocols across cryptographic tasks. Like SCVM [32], Viaduct can synthesize “hybrid” programs that perform computations locally, replicated between hosts, or under MPC. This is impossible in the simple two-point label model that many MPC compilers [3, 33] use, which only distinguish between public (low) and secret (high) information. Viaduct also does not fix the number of hosts in a program (unlike [32–34]), nor fix compiling programs only under a semi-honest or malicious setting (unlike [31–33, 37, 40, 46]).

**Program Partitioning.** Another line of related work [19, 20, 51, 52] describes distributed computations using sequential programs and captures security requirements using information-flow labels. The Jif/split compiler [51, 52] synthesizes simple cryptographic primitives such as cryptographic commitments to satisfy security constraints that would otherwise be impossible without relying on trusted principals. Unlike Viaduct, Jif/split is not extensible to new protocols. Later work [19, 20] proves computational soundness for a similar system under a strong attacker that controls the network and some of the hosts. However, this work does not support replicating computations (only data replication is supported), or the other protocols that Viaduct supports.

9 Conclusion

The prototype implementation of the Viaduct compiler compiles high-level, security-typed programs into efficient distributed programs that employ a variety cryptographic mechanisms to ensure security. And the compiler is agnostic to the set of available protocols, making it easily extensible.

Promising avenues for future work remain. The label model could be extended with availability policies [53], guiding selection of fault-tolerant protocols like quorum replication [54] and MPC with guaranteed output delivery [26]. A more full-fledged implementation of Viaduct could support executing code on trusted execution environments like hardware enclaves [25, 29, 35], the use of special-purpose protocols like private set intersection [12, 38] and Oblivious RAM [44], and the incorporation of a more detailed and accurate cost model [28].

Finally, a full correctness proof for the Viaduct compiler would be a significant research achievement, bridging security notions defined by the programming-languages and cryptography communities. One can see Viaduct source programs as *ideal functionalities* and the distributed programs generated by the compiler as *hybrid protocols* using ideal functionalities implemented by cryptographic mechanisms. The conjectured correctness statement for Viaduct is a simulation proof in the Universal Composability (UC) framework [9], relating a Viaduct source program to the distributed implementation generated by the compiler.

Acknowledgments

This work was supported by the National Science Foundation under grant CNS-1704788, and by the IARPA HECTOR program under a subcontract from IBM. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any of these funders.

We thank our shepherd Mike Hicks and our anonymous reviewers for their insightful suggestions. We also thank Alexa van Hattum, Tobias Kappe, Ralph Recto, and Drew Zagieboylo for feedback during the drafting of this paper.
References


