# Town Crier:
# An Authenticated Data Feed for Smart Contracts


## Abstract

Smart contracts are programs that execute autonomously on blockchains. Many of their envisioned uses require them to consume data from outside the blockchain. (For example, a financial instrument might rely on stock prices.) Trustworthy *data feeds* that can support data requests by smart contracts will thus be critical to any smart contract system.

We present an authenticated data feed system called Town Crier (TC). TC builds on the observation that many web sites, such as major news and finance sites, already serve as trusted data sources for non-blockchain uses. TC acts as a bridge between such servers and smart contract systems. It uses trusted hardware to authenticate and scrape data from HTTPS-enabled websites and to generate trustworthy data for relying smart contracts. It also includes a range of advanced features, such as support for private data requests, which involve decryption and evaluation of request ciphertext within TC's hardware.

We describe the TC architecture, its underlying trust model, and its applications, and report on an implementation that uses the newly released Ingel SGX software development kit and furnishes data for the smart-contract system Ethereum. Finally, we present formal proofs of the security of TC, including correct handling of payment in Ethereum. We will soon be launching TC as an online public service.


## 1 Introduction

Smart contracts are computer programs that autonomously facilitate or execute a contract. For decades, they have been envisioned as a means to render existing contract law more precise, pervasive, and efficiently executable through logical specification of legal agreements. Szabo, who popularized the term "smart contact" in a seminal 1997 essay [**?**], gave as an example a smart contract that enforces car loan payments. If the owner of the car fails to make a timely payment, a smart contract could programmatically revoke physical access and return control of the car to the bank.

Cryptocurrencies such as Bitcoin [**?**] have provided key technical provisions for decentralized smart contracts: Direct control of money by programs and fair, automated execution of computations through the decentralized consensus mechanisms underlying blockchains. The recently launched Ethereum furthermore supports Turing-complete code and thus fully expressive, self-enforcing smart contracts, a big step toward the vision of researchers and proponents.

As Szabo's example shows, however, the most compelling applications of smart contracts require not just blockchain code, but access to data about real-world state and events. Similarly, financial contracts and derivatives, key applications for Ethereum [**?**, **?**], rely on data about financial markets, such as equity prices.

*Data feeds* (a.k.a. "oracles"), contracts on the blockchain that serve data requests by other contracts [**?**, **?**], are intended to meet this need. A few data feeds exist for Ethereum today, but provide no assurance of trustworthy data beyond the reputation of their operators (who are typically individuals or small entities), even if their data originates with trustworthy sources. Of course, there exist reputable websites that serve data for non-blockchain applications and sometimes use HTTPS, enabling source authentication of served data. Smart contracts, though, lack network access and thus cannot directly access such data. The lack of a substantive ecosystem of trustworthy data feeds thus remains an oft-cited, critical obstacle to the evolution of in Ethereum and smart contracts in general [**?**].

**Town Crier.** We introduce a system called *Town Crier* (TC) that provides an *authenticated data feed* (ADF) for smart contracts. TC acts as a high-trust bridge between existing HTTPS-enabled data websites and the Ethereum blockchain. It retrieves website data and serves it to relying contracts on the blockchain as concise, contract-consumable pieces of data (e.g., stock quotes) called *datagrams*. TC makes use of Software Guard Instructions (SGX), Intel's recently released trusted hardware capability. It executes its core functionality as a trusted piece of code in an SGX *enclave*, which protects against malicious processes and the OS and can *attest* (prove) to a remote client that the client is interacting with a legitimate, SGX-backed instance of the TC code.

Through a smart-contract front end, Town Crier respond to requests by contracts on the blockchain with attestations of the following form:

> "Datagram $X$ specified by parameters params is served by an HTTPS-enabled website $Y$ during a specified time frame $T$."

A relying contract can be assured of the correctness of $X$ in such a datagram if it trusts the security of SGX, the

(published) TC code, and the validity of source data in the specified interval of time.

**Contributions of TC.** We highlight several important contributions in our design of TC:

*Fully functional TC implementation, with pending open source and launch.* We designed and implemented Town Crier as a complete, end-to-end system that offers formal security guarantees at the cryptographic protocol level. Aiming beyond an advance in academic research, we plan to launch Town Crier as an open-source, production service atop Ethereum in the near future. Our launch of TC awaits only availability of the Intel Attestation Service (IAS), which is expected to occur soon. In its initial form, Town Crier be a free service for smart contract users, requiring users only to defray the (small) gas cost of invoking TC on the Ethereum blockchain.

*Formal security analysis.* Formal security is vitally important in a data feed, as smart contracts execute in an adversarial environment where contractual parties can reap financial gains by subverting smart contracts and or services they rely upon. Legal recourse is often impractical precisely because smart contracts beneficially enable micro-services without costly legal setup and enforcement, as well as contracts between arbitrary, pseudonymous users. We thus adopt a rigorous, principled approach to the design of Town Crier by formally defining and ensuring:

- *Authenticity*: A datagram $X$ returned to a requesting contract is guaranteed to truly reflect the data served by specified website $Y$ in time interval $T$ with the requester's specified parameters params.
- *Gas neutrality (zero TC loss).* Assuming that *TC* executes honestly, it does not lose money (cannot suffer resource depletion) even when accessed by arbitrarily malicious blockchain contracts and users.
- *Fair expenditure (bounded requester loss).* Even when all other users, contracts, and TC itself act maliciously, an honest requester will never pay more than a tiny amount beyond what is required for valid computation executed for that request—whether or not a datagram is delivered.

To obtain the above formal guarantees, we rely on the formal modeling of blockchains proposed by Kosba et al. [**?**] and the formal abstraction for SGX proposed by Shi et al. [**?**] Our analysis of TC reveals interesting challenges and technical subtleties, e.g., a subtle gap between the formal blockchain model of Kosba et al. [**?**] and Ethereum's instantiation that proves important in formal reasoning about TC's gas handling.

*Robustness to component compromise.* TC minimizes the Trusted Computing Base (TCB) of its trusted code in the SGX enclave. It thus offers a basic security model in which a user need only trust SGX itself and a designated data source (website). As an additional feature, TC can hedge against the risk of a single SGX instance or website compromise by supporting various modes of majority voting, among multiple websites offering the same piece of data (e.g., stock price), and/or multiple, possibly geographically dispersed SGX instances.

*Private and custom datagrams.* To meet the potentially complex confidentiality concerns that can arise in the broad array of smart contracts enabled by TC, TC's trusted enclave code is instrumented to ingest confidential user data (encrypted under a TC public key). It can thereby support *private* datagram requests, with encrypted parameters, and *custom* datagram requests, which securely access the online resources of requesters (e.g., online accounts) using encrypted credentials.

**Applications.** Thanks to the above key contributions, we believe that TC can spur deployment of a rich spectrum of smart contracts that are hard to realize in the existing Ethereum ecosystem. We present three examples that showcase TC's capabilities and demonstrate its end-to-end use: (1) A financial derivative (cash-settled put option) that consumes stock ticker data; (2) A flight insurance contract that relies on private data requests about flight cancellations; and (3) A contract for sale of virtual goods and online games (via Steam Marketplace) for ether, the Ethereum currency, using custom data requests to access online user accounts. We experimentally measure response times for associated datagram requests ranging from 192-1309 ms, depending on the datagram type. These times are significantly less than an Ethereum block interval, and suggest that a few SGX-enabled hosts can support TC data feed rates well beyond the global transaction rate of a modern decentralized blockchain.

***Organization:*** We present basic technical background for TC (Section **??**), followed by an architectural description (Section **??**), a basic set of protocols (Section **??**), a discussion of implementation details (Section **??**), and formal security analysis (Section **??**). We present three example applications (Section **??**) and use these applications as the basis for performance evaluations of TC (Section **??**). We present related work (Section **??**) and conclude (Section **??**) with a brief discussion of future work. The paper appendix includes formalism and future directions omitted from the paper body.

## 2 Background

In this section, we provide basic background respectively on the main technologies TC incorporates, namely SGX, TLS / HTTPS, and smart contracts.

**SGX.** Intel's Software Guard Extensions (SGX) [**?**, **?**, **?**, **?**, **?**] is a set of new instructions that confer hardware protections on user-level code. SGX enables a process to execute in a protected address space known as an *enclave*, which protects the confidentiality and integrity of the process from other software on the same host, including the operating system, and certain forms of hardware attack.

A enclave process cannot make system calls, but can read and write memory outside the enclave region. Thus isolated execution in SGX may be viewed in terms of an ideal model in which a process is guaranteed to execute correctly and with perfect confidentiality, but relies on a (potentially malicious) operating system for network and file-system access.[1]

SGX allows a remote system to verify the software in an enclave and communicate securely with it. When an enclave is created, the CPU produces a hash of its initial state known as a *measurement*. The software in the enclave may at a later time request a report, which includes a measurement and supplementary data provided by the process, such as a public key. The report may be digitally signed using a hardware-protected key to produce a proof that the measured software is running in an SGX-protected enclave. This proof, known as a *quote*, may be verified by a remote system, while the process-provided public key can be used by the remote system to establish a secure channel with the enclave or verify signed data it emits. We use the generic term *attestation* to refer to a quote, and denote it by `att`. We assume that a trustworthy measurement of the code for the enclave component of TC is available to any client that wishes to verify an attestation. SGX signs quotes using a *group signature* scheme called EPID [**?**]. This choice of primitive is significant in our design of Town Crier, as EPID is a proprietary signature scheme not supported in Ethereum.

SGX additionally provides a trusted time source via the function `sgx_get_trusted_time`. On invoking this function, an enclave obtains a measure of time relative to a reference point indexed by a nonce. A reference point remains stable, but SGX does not provide a source of absolute or wall-clock time, a limitation that we must work around in TC.

**TLS / HTTPS.** We assume basic familiarity by readers with TLS and HTTPS (HTTP over TLS). As we explain later, TC exploits an important feature of HTTPS, namely that it can be partitioned into interoperable layers: An HTTP layer interacting with web servers, a TLS layer handling handshakes and secure communication, and a TCP layer providing reliable data stream.

[1]This model is a simplification: SGX is known to expose some internal enclave state to the OS [**?**]. Our basic security model for TC assumes ideal isolated execution, but again, TC can also be distributed across multiple SGX instances as a hedge against compromise.

**Smart contracts.** While TC can in principle support any smart-contract system, we focus in this paper on its use in Ethereum, whose use we now explain. For further details, see [**?**, **?**]

A smart contract in Ethereum is represented as what is called a *contract account*, endowed with code, a currency balance, and persistent memory in the form of a key/value store. A contract accepts messages as inputs to any of a number of designated functions. These entry points, determined by the contract creator, represent the API of the contract. Once created, a contract executes autonomously; it persists indefinitely, with even its creator unable to modify its code.[2] Contract code executes in response to receipt of a *message* from another contract or a *transaction* from a non-contract (*externally owned*) account, informally what we call a *wallet*. Thus, contract execution is always initiated by a transaction. Informally, a contract only executes when "poked," and poking progresses through a sequence of entry points until no further message passing occurs (or a shortfall in gas occurs, as explained below). The "poking" model aside, as a simple abstraction, a smart contract may be viewed as an *autonomous agent* on the blockchain.

Ethereum has its own associated cryptocurrency called *ether*. (At the time of writing, 1 ether has a market value of a little more than $5 U.S. [**?**].) To prevent Denial-of-Service (DoS) attacks, inadvertent infinite looping within contracts, and generally to control network resource expenditure, Ethereum allows ether-based purchase of a resource called *gas* to power contracts. Every operation, including sending data, executing computation, and storing data, has a fixed gas cost. Transactions and messages include a parameter (`GASLIMIT`) specifying a bound on the amount of gas expended by the computations they initiate. When a function call is made, the child function expends gas from the same source as the parent function. Should a function fail to complete due to a gas shortfall, it is aborted and any state changes induced by the partial computation are rolled back to their pre-call state; previous computations on the call path, though, are retained.

Along with a `GASLIMIT`, a transaction specifies a `GASPRICE`, the maximum amount in ether that the transaction is willing to pay per unit of gas. The transaction thus succeeds only if the initiating account has a balance of `GASLIMIT` $\times$ `GASPRICE` ether and `GASPRICE` is high enough to be accepted by the system (miner).

The management of gas, as we show in our design of Town Crier can be delicate. Without careful construction, TC's smart contract front end on the Ethereum blockchain can be caused by an attacker to exhaust the ether used to power the delivery of datagrams.

Finally, we note that transactions in Ethereum are dig-

[2]There is one exception: A special opcode `suicide` wipes code from a contract account.

itally signed for a wallet using ECDSA on the curve Secp256k1 and the hash function SHA3-256.

## 3 TC Architecture and Security Model

Town Crier includes three main components: The TC Contract ($\mathcal{C}_{TC}$), the Enclave (whose code is denoted by $\mathsf{prog}_{\mathsf{encl}}$), and the Relay ($\mathcal{R}$). The Enclave and Relay reside on the TC server, while the TC Contract resides on the blockchain. We denote a smart contract making use of the Town Crier service as a *requester* or *relying* contract; we denote this contract by $\mathcal{C}_U$, and its (off-chain) owner as a *client* or *user*. A *data source*, or *source* for short, is an online server (running HTTPS) that provides data which TC draws on to compose datagrams.

An architectural schematic of TC showing its interaction with external entities is given in Figure **??**.

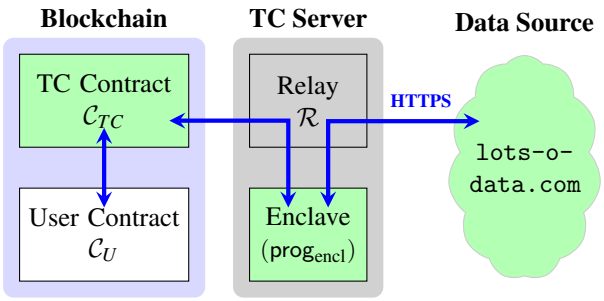


Figure 1: **Basic Town Crier architecture.** Trusted components are depicted in green.

**The TC Contract $\mathcal{C}_{TC}$.** The TC Contract is a smart contract that acts as the blockchain front end for TC. It is designed to present a simple API to a relying contract $\mathcal{C}_U$ for its requests to TC. Very simply, $\mathcal{C}_{TC}$ accepts datagram requests from a requester $\mathcal{C}_U$ and returns corresponding datagrams from TC. Additionally, $\mathcal{C}_{TC}$ manages TC monetary resources, namely ether and gas.

**The Enclave.** We refer to an instance of the TC code running in an SGX enclave simply as the Enclave and denote the code itself by $\mathsf{prog}_{\mathsf{encl}}$. In TC, the Enclave ingests and fulfills datagram requests from the blockchain. To obtain the data for inclusion in datagrams, it queries external data sources, specifically HTTPS-enabled internet services. It returns a datagram to a requesting contract $\mathcal{C}_U$ as a digitally signed blockchain message. Under our basic security model for SGX, network functions aside, the Enclave runs in complete isolation from an adversarial OS as well as other process on the host.

**The Relay $\mathcal{R}$.** As an SGX enclave process, the Enclave lacks direct network access. Thus the Relay handles bidirectional network traffic on behalf of the Enclave. Specifically, the Relay provides network connectivity from the Enclave to three different types of entities:

1. *The Blockchain (the Ethereum system):* The Relay scrapes the blockchain in order to monitor the state of the TC Contract $\mathcal{C}_{TC}$. In this way, it performs implicit message passing from $\mathcal{C}_{TC}$ to the Enclave, as neither component itself has network connectivity. Additionally, the Relay places messages emitted from the Enclave (datagrams) on the blockchain.

2. *Clients:* The Relay runs a web server to handle off-chain service requests from clients, specifically, requests for attestations from the Enclave. As we soon explain, an attestation provides a unique public key for the Enclave instance to the client and proves that the Enclave is executing correct code in an enclave and that its clock is correct in terms of absolute (wall-clock time). A client that successfully verifies an attestation can then safely create a relying contract $\mathcal{C}_U$ that uses the TC.

3. *Data sources:* The Relay relays traffic to and from data sources (HTTPS-enabled websites) queried by the Enclave.

The Relay is an ordinary user-space application. It does not benefit from integrity protection by SGX and thus, unlike the Enclave, can be subverted by an adversarial OS on the TC server, causing network delays or failures. A key design aim of TC, however, is that Relay should be unable to cause incorrect datagrams to be produced or users to lose fees paid to TC for datagrams (although they may lose gas used to fuel their requests). As we shall show, in general the Relay *can only mount denial-of-service attacks against TC*.

**Security model.** Here we give a brief overview of our security model for TC, providing more details in later sections. We assume the following:

- *The TC Contract.* $\mathcal{C}_{TC}$ is globally visible on the blockchain and its source code is published for clients. Thus we assume that $\mathcal{C}_{TC}$ behaves honestly.
- *Data sources.* We assume that clients trust the data sources from which they obtain TC datagrams. We also assume that these sources are stable, i.e., yield consistent datagrams, during a requester's specified time interval $T$. (Requests are generally time-invariant, e.g., for a stock price at a particular time.)
- *Enclave security.* We make three assumptions: (1) The Enclave behaves honestly, i.e., $\mathsf{prog}_{\mathsf{encl}}$, whose source code is published for clients, correctly executes the TC protocol; (2) For an Enclave-generated keypair $(\mathsf{sk}_{TC}, \mathsf{pk}_{TC})$, the private key $\mathsf{sk}_{TC}$ is known only to the Enclave; and (3) The Enclave has an accurate (internal)

real-time clock. We explain below how we use SGX to achieve these properties.

- *Blockchain communication.* Transaction and message sources are authenticable, i.e., a transaction $m$ sent from an account / wallet $\mathcal{W}_X$ (or message $m$ from contract $\mathcal{C}_X$) is identified by the receiving account as originating from $X$. Transactions and messages are integrity protected (as they are digitally signed by the sender), but not confidential.
- *Network communication.* The Relay (and other untrusted components of the TC server) can tamper with or delay communications to and from the Enclave. (As we explain in our SGX security model, the Relay cannot otherwise observe or alter the behavior of the Enclave.) Thus the Relay is subsumed by an adversary that controls the network.

# 4 Basic TC Protocol

We now describe the operation of TC at the protocol level. The basic protocol is conceptually simple: A user contract $\mathcal{C}_U$ requests a datagram from the TC Contract $\mathcal{C}_{TC}$. $\mathcal{C}_{TC}$ forwards the request to $\mathsf{prog}_{\mathrm{encl}}$ and then returns the request to $\mathcal{C}_U$. There are many details, however, relating to message contents and protection and the need to connect the off-chain parts of TC with the blockchain.

First, we give a brief protocol overview. Then we enumerate the data flows in TC. Finally, we provide a component-level view of the protocol by specifying the operation of the TC Contract, Relay, and Enclave. We present these as ideal functionalities, inspired by the universal-composability (UC) framework, in order to abstract away implementation details and as a springboard for formal proofs of security. We omit details in this section on how payment is incorporated into TC, deferring this delicate aspect of the system design to Section **??**.

## 4.1 Datagram lifecycle

The lifecycle of a datagram may be briefly summarized in the following steps:

- **Initiate request.** $\mathcal{C}_U$ sends a datagram request to $\mathcal{C}_{TC}$ on the blockchain.
- **Monitor and relay.** The Relay monitors $\mathcal{C}_{TC}$ and relays any incoming datagram request with parameters $\mathsf{params}$ to the Enclave.
- **Securely fetch feed.** To process the request specified in $\mathsf{params}$, the Enclave contacts a data source via HTTPS and obtains the requested datagram. It forwards the datagram via the Relay to $\mathcal{C}_{TC}$.
- **Return datagram.** $\mathcal{C}_{TC}$ returns the datagram to $\mathcal{C}_U$.

We now make this data flow more precise.

## 4.2 Data flows

A datagram request by $\mathcal{C}_U$ takes the form of a message $m_1 = (\mathsf{params}, \mathsf{callback})$ to $\mathcal{C}_{TC}$ on the blockchain. $\mathsf{params}$ specifies the requested datagram, e.g., $\mathsf{params} := (\mathsf{url}, \mathsf{spec}, T)$, where $\mathsf{url}$ is the target data source and $\mathsf{spec}$ specifies content of a the datagram to be retrieved (e.g., a stock ticker at a particular time), while $T$ specifies the delivery time for the datagram (initiated by scraping of the data source). The parameter $\mathsf{callback}$ in $m_1$ indicates the entry point in $\mathcal{C}_U$ to which the datagram is to be returned. (In principle, $\mathsf{callback}$ could point to a different contract, but TC does not yet adopt this generalization.)

$\mathcal{C}_{TC}$ forwards $m_2 = (\mathsf{id}, \mathsf{params})$ to the Enclave. It receives in return a return message $m_3 = (\mathsf{id}, \mathsf{params}, \mathsf{data})$ from the $TC$ service, where $\mathsf{data}$ is the datagram, i.e., contains the data (e.g., the desired stock ticker price). $\mathcal{C}_{TC}$ checks the consistency of $\mathsf{params}$ on the incoming and outgoing messages, and if they match forwards $\mathsf{data}$ to the entry point $\mathsf{callback}$ in $\mathcal{C}_U$ in message $m_4$. Finally, $\mathcal{C}_U$ uses $\mathsf{id}$ in $m_4$ to match the returned datagram to its original request.

For simplicity here, we assume that $\mathcal{C}_U$ makes a one-time datagram request. Thus it can trivially match $m_4$ with $m_1$. Our full protocol contains an optimization by which $\mathcal{C}_{TC}$ assigns $\mathsf{id}$ to $m_1$ (and sends $\mathsf{id}$ to $\mathcal{C}_U$), creating a consistent, trustworthy identifier for all data flows, and enabling straightforward handling of multiple datagram requests from the same instance of $\mathcal{C}_U$.

Fig. **??** shows the data flows involved in processing a datagram request. For simplicity, the figure omits the Relay, which is only responsible for data passing.

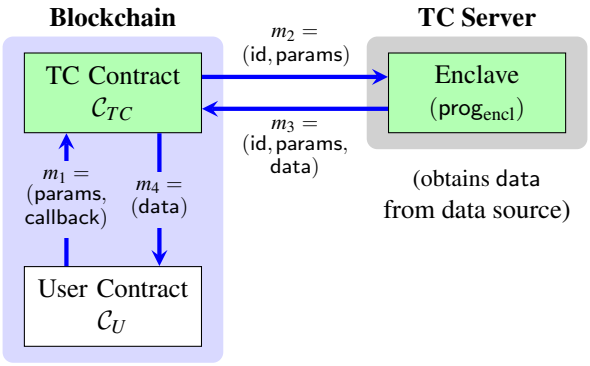


Figure 2: **Data flows in datagram processing.**

Digital signatures are needed to authenticate messages, such as $m_3$, entering the blockchain from an external source. We let $(\mathsf{sk}_{TC}, \mathsf{pk}_{TC})$ denote the private / public keypair associated with the Enclave for such message authentication. For simplicity, Fig. **??** assumes that the Enclave can send signed messages directly to $\mathcal{C}_{TC}$. We explain shortly how Ethereum requires a layer of in-

direction such that TC sends $m_3$ as a transcation via an Ethereum wallet $\mathcal{W}_{TC}$.

## 4.3 Use of SGX

Let $\mathsf{prog}_{\text{encl}}$ represent the code for Enclave, which we presume is trusted by all system participants. Our protocols in TC rely on the ability of SGX to attest to execution of an instance of $\mathsf{prog}_{\text{encl}}$ and bind a public key $\mathsf{pk}_{TC}$ to this instance. Here we briefly explain how we achieve these goals. First, we present a model that abstracts away implementation details in SGX, helping simplify our protocol presentation and later, our security proofs. We then explain how SGX attestation is used to authenticate datagrams served by $\mathcal{C}_{TC}$, namely through binding of $\mathsf{pk}_{TC}$ to an Ethereum wallet on the blockchain. Finally, we explain how we use the clock in SGX. Our discussion draws on formalism for SGX from [**?**].

**Formal model and notation.** We adopt a formal abstraction of Intel SGX proposed by Shi et al. [**?**]. Following the UC and GUC paradigms [**?**, **?**, **?**], Shi et al. propose to abstract away the details of SGX implementation, and instead view SGX as a third party trusted for both confidentiality and integrity. Specifically, we use a global UC functionality $\mathcal{F}_{\text{sgx}}(\Sigma_{\text{sgx}})[\mathsf{prog}_{\text{encl}}, \mathcal{R}]$ to denote (an instance of) an SGX functionality parametrized by a (group) signature scheme $\Sigma_{\text{sgx}}$. Here $\mathsf{prog}_{\text{encl}}$ denotes the SGX enclave program and $\mathcal{R}$ the physical SGX host (which we assume for simplicity is the same as that for the TC Relay). Upon initialization, $\mathcal{F}_{\text{sgx}}$ runs $\mathsf{outp} := \mathsf{prog}_{\text{encl}}.\textbf{Initialize}$ and attests to the code of $\mathsf{prog}_{\text{encl}}$ as well as $\mathsf{outp}$. Upon a resume call with $\mathsf{params}$, $\mathcal{F}_{\text{sgx}}$ runs and outputs the result of $\mathsf{prog}_{\text{encl}}.\textbf{Resume}(\mathsf{params})$. Further formalism for $\mathcal{F}_{\text{sgx}}$ is given in Appendix **??**.

**Binding $\mathsf{prog}_{\text{encl}}$ to Ethereum wallet $\mathcal{W}_{TC}$.** Information can only be inserted into the blockchain in Ethereum as a transaction from a wallet. Thus, the only way the Relay can relay messages from the Enclave to $\mathcal{C}_{TC}$ is through a wallet $\mathcal{W}_{TC}$. Since the Relay may corrupt messages, however, it is critical that they be authenticated by the Enclave. Since Ethereum itself already verifies signatures on transactions from externally owned accounts (i.e., users interact with the Ethereum blockchain through an authenticated channel), TC uses a trick to *piggyback verification of enclave signatures on top of Ethereum's already existing transaction signature verification mechanism*.

Very simply, the Enclave creates $\mathcal{W}_{TC}$ with the public key $\mathsf{pk}_{TC}$.

To make this idea work fully, the public key $\mathsf{pk}_{TC}$ must be hardcoded into $\mathcal{C}_{TC}$. A client creating or relying on a contract that uses $\mathcal{C}_{TC}$ is responsible for making sure that this hardcoded $\mathsf{pk}_{TC}$ has an appropriate SGX attestation

**$\mathcal{F}_{\text{sgx}}[\mathsf{prog}_{\text{encl}}, \mathcal{R}]$: abstraction for SGX**

***Hardcoded***: $\mathsf{sk}_{\text{sgx}}$

***Assume***: $\mathsf{prog}_{\text{encl}}$ has entry points **Initialize** and **Resume**

**Initialize:**

On receive "initialize" from $\mathcal{R}$:
  Let $\mathsf{outp} := \mathsf{prog}_{\text{encl}}.\textbf{Initalize}()$
  $\sigma_{\text{att}} := \Sigma_{\text{sgx}}.\mathsf{Sign}(\mathsf{sk}_{\text{sgx}}, (\mathsf{prog}_{\text{encl}}, \mathsf{outp}))$
    *// models EPID sig.*
  Output $(\mathsf{outp}, \sigma_{\text{att}})$

**Resume:**

On receive ("resume", $\mathsf{params}$) from $\mathcal{R}$:
  Let $\mathsf{outp} := \mathsf{prog}_{\text{encl}}.\textbf{Resume}(\mathsf{params})$
  Output $\mathsf{outp}$

Figure 3: Formal abstraction for SGX execution capturing a subset of SGX features sufficient for implementation of TC.

before interacting with the $\mathcal{C}_{TC}$ blockchain contract. Let $\mathsf{Verify}$ denote a verification algorithm for EPID signatures. Fig. **??** gives the protocol for a client to check that $\mathcal{C}_{TC}$ is backed by a valid Enclave instance. (We omit modeling here of IAS online revocation checks.)

In summary, then, we may assume in our protocol specifications that *relying clients have verified an attestation for Enclave and thus that datagram responses sent from $\mathcal{W}_{TC}$ to $\mathcal{C}_{TC}$ are trusted to originate from* $\mathsf{prog}_{encl}$.

**User: offline verification of SGX attestation**

***Inputs***: $\mathsf{pk}_{\text{sgx}}$, $\mathsf{pk}_{TC}$, $\mathsf{prog}_{\text{encl}}$, $\sigma_{\text{att}}$

**Verify:**

Assert $\mathsf{prog}_{\text{encl}}$ is the expected enclave code
Assert $\Sigma_{\text{sgx}}.\mathsf{Verify}(\mathsf{pk}_{\text{sgx}}, \sigma_{\text{att}}, (\mathsf{prog}_{\text{encl}}, \mathsf{pk}_{TC}))$
Assert $\mathcal{C}_{TC}$ is correct and parametrized w/ $\mathsf{pk}_{TC}$
*// now okay to rely on $\mathcal{C}_{TC}$*

Figure 4: A client checks an SGX attestation on the enclave's code $\mathsf{prog}_{\text{encl}}$ and public key $\mathsf{pk}_{TC}$; the client checks that $\mathsf{pk}_{TC}$ is hardcoded into TC blockchain contract $\mathcal{C}_{TC}$ before using $\mathcal{C}_{TC}$.

**SGX Clock.** As noted above, the trusted clock for SGX provides only relative time with respect to a reference point, not absolute time. Thus, when initialized, the Enclave is provided with the current wall-clock time by a trusted source, e.g., the Relay (under a trust-on-first-use model). In the current implementation of TC, clients may request and verify in real time a fresh timestamp, signed by the Enclave under $\mathsf{pk}_{TC}$, via a web interface in the Re-

lay. Thus, a client can determine the absolute clock time of Enclave to within a high degree of accuracy, bounded by the round-trip time of its attestation request plus the attestation verification time–on the order of hundreds of milliseconds in a wide-area network []. A high degree of accuracy is potentially useful for some applications but only loose accuracy is not required for most. Ethereum targets a block interval of 12s and the clock serves in TC primarily to: (1) Schedule connections to data sources and (2) To check TLS certificates for expiration when establishing HTTPS connections. For simplicity, we just assume in our protocol specifications that the Enclave clock provides accurate wall-clock time (in the canonical format of seconds since the Unix epoch January 1, 1970 00:00 UTC).

We let clock() denote measurement of the SGX clock from within the enclave, expressed as the current absolute (wall-clock) time.

## 4.4 A payment-free basic protocol

For simplicity, we first specify a payment-free version of our basic protocol, i.e., one that does not include gas or fees. Later, in our implementation discussion, we explain how we handle these two resources, and we prove payment-related properties in the paper appendix. For simplicity, we assume a single instance of $\mathsf{prog}_{\text{encl}}$, although our architecture could scale up to multiple enclave and even server instances. To show messages corresponding to those in Fig. **??**, we use the label (**msg.** $m_i$)

**The Requester Contract $\mathcal{C}_U$.** The requester contract $\mathcal{C}_U$ sends to the TC Contract $\mathcal{C}_{TC}$ a request of the form $(\mathsf{params}, \mathsf{callback})$.

**The TC Contract $\mathcal{C}_{TC}$.** The TC Contract, as noted above, accepts a datagram request from $\mathcal{C}_U$, assigns a unique id to each request, and records the request. Our Town Crier Relay $\mathcal{R}$ will monitor requests received by $\mathcal{C}_{TC}$, and forwards the requests to an SGX enclave. When $\mathcal{C}_{TC}$ obtains a valid response from $\mathcal{W}_{TC}$, it sends the resulting datagram data to the entry point callback specified by the requesting contract $\mathcal{C}_U$. As explained above, because the response (msg. $m_2$) comes from $\mathcal{W}_{TC}$, the blockchain automatically verifies that the response is correctly signed under $\mathsf{prog}_{\text{encl}}$'s key $\mathsf{pk}_{TC}$, and $\mathcal{C}_{TC}$ need not verify the signature explicitly. TC does, however, have a subtle security requirement. Specifically, for a given datagram request id, $\mathcal{C}_{TC}$ must verify that $\mathsf{params}' = \mathsf{params}$, where $\mathsf{params}'$ is the digitally signed message produced by $\mathsf{prog}_{\text{encl}}$ and params is the locally stored parameters. The check is necessary to prevent $\mathcal{R}$ from corrupting datagram requests passed by $\mathcal{C}_{TC}$ (which, as a public function, has no means of digitally signing requests).

$\mathcal{C}_{TC}$ is specified in Fig. **??**. Here, Call denotes a call to a contact entry point, while Return denotes a message returned by the blockchain in response to a function call.

**Program for Town Crier blockchain contract $\mathcal{C}_{TC}$**

**Initialize:** Counter := 0

**Request:** On recv $(\mathsf{params}, \mathsf{callback})$ from some $\mathcal{C}_U$:
id:= Counter; Counter := Counter + 1
Record $(\mathsf{id}, \mathsf{params}, \mathsf{callback})$ // **msg.** $m_1$

**Deliver:** On recv $(\mathsf{id}, \mathsf{params}, \mathsf{data})$ from $\mathcal{W}_{TC}$:
Retrieve recorded $(\mathsf{id}, \mathsf{params}', \mathsf{callback})$
Assert $\mathsf{params} = \mathsf{params}'$
Call $\mathsf{callback}(\mathsf{data})$ // **msg.** $m_4$

Figure 5: The Town Crier TC Contract $\mathcal{C}_{TC}$.

**The Enclave $\mathsf{prog}_{\text{encl}}$.** When initialized through an `init` call, $\mathsf{prog}_{\text{encl}}$ ingests the current wall-clock time; by storing this time and setting a clock reference point, it calibrates its absolute clock. It generates an ECDSA keypair $(\mathsf{pk}_{TC}, \mathsf{sk}_{TC})$ (parameterized as in Ethereum), where $\mathsf{pk}_{TC}$ is bound to the $\mathsf{prog}_{\text{encl}}$ instance through insertion into attestations.

On input `resume` $(\mathsf{id}, \mathsf{params})$, $\mathsf{prog}_{\text{encl}}$ contacts the data source specified by params via HTTPS and checks that the corresponding certificate cert is valid. (We discuss certificate checking in Section **??**.) Then $\mathsf{prog}_{\text{encl}}$ fetches the requested datagram and returns it to $\mathcal{R}$ along with params and id, all digitally signed with $\mathsf{sk}_{TC}$. Fig. **??** shows the protocol for $\mathsf{prog}_{\text{encl}}$.

**Program for Town Crier Enclave ($\mathsf{prog}_{\text{encl}}$)**

**Initialize**: On recv (`init`):
$(\mathsf{pk}_{TC}, \mathsf{sk}_{TC}) := \Sigma.\mathsf{KeyGen}(1^\lambda)$
Output $\mathsf{pk}_{TC}$
*/* Subroutine call from $\mathcal{F}_{sgx}$, which attests to $\mathsf{prog}_{encl}$ and $\mathsf{pk}_{TC}$. See Figure* **??***.*/*

**Resume:** On recv (`resume`, $(\mathsf{id}, \mathsf{params})$)
Parse params as $(\mathsf{url}, \mathsf{spec}, T)$:
Assert $\mathsf{clock}() \geq T.\mathsf{min}$
Contact url via HTTPS, obtaining cert
Verify cert is valid for time clock()
Obtain webpage $w$ from url
Assert $\mathsf{clock}() \leq T.\mathsf{max}$
Parse $w$ to extract data with specification spec
$\sigma := \Sigma.\mathsf{Sign}(\mathsf{sk}_{TC}, (\mathsf{id}, \mathsf{params}, \mathsf{data}))$
Output $((\mathsf{id}, \mathsf{params}, \mathsf{data}), \sigma)$

Figure 6: The Town Crier Enclave $\mathsf{prog}_{\text{encl}}$.

**The Relay $\mathcal{R}$.** As noted in Section **??**, $\mathcal{R}$ performs

distinct three forms of data passing, which we specify more precisely here. It scrapes the blockchain, monitoring $\mathcal{C}_{TC}$ for new datagram requests $(\mathsf{id}, \mathsf{params})$. It boots the $\mathsf{prog}_{\mathsf{encl}}$ with an `init` call and handles incoming requests by invoking $\mathsf{prog}_{\mathsf{encl}}$ with a `resume` call. Finally, it forwards datagram responses from $\mathsf{prog}_{\mathsf{encl}}$ to the blockchain; recall that it inserts transactis to the blockchain and thus forward responses as $\mathcal{W}_{TC}$. The program for $\mathcal{R}$ is shown in Fig. **??**. Here, the function AuthSend inserts a transaction to blockchain ("as $\mathcal{W}_{TC}$" means the transaction is signed with $\mathsf{sk}_{TC}$).

**Program for Town Crier Relay $\mathcal{R}$**

**Initialize**:
Send `init` to $\mathcal{F}_{\text{sgx}}[\mathsf{prog}_{\mathsf{encl}}, \mathcal{R}]$
On recv $(\mathsf{pk}_{TC}, \sigma_{\text{att}})$ from $\mathcal{F}_{\text{sgx}}[\mathsf{prog}_{\mathsf{encl}}, \mathcal{R}]$:
  Publish $(\mathsf{pk}_{TC}, \sigma_{\text{att}})$

**Handle**$(\mathsf{id}, \mathsf{params})$:
Parse params as $(\_, \_, T)$
Wait until $\mathsf{clock}() \geq T.\mathsf{min}$
Send $(\texttt{resume}, (\mathsf{id}, \mathsf{params}))$ to $\mathcal{F}_{\text{sgx}}[\mathsf{prog}_{\mathsf{encl}}, \mathcal{R}]$
On recv $((\mathsf{id}, \mathsf{params}, \mathsf{data}), \sigma)$ from $\mathcal{F}_{\text{sgx}}[\mathsf{prog}_{\mathsf{encl}}, \mathcal{R}]$:
  AuthSend $(\mathsf{id}, \mathsf{params}, \mathsf{data})$ to $\mathcal{C}_{TC}$ as $\mathcal{W}_{TC}$
  *//send out* **msg.** $m_3$

**Main**:
Loop Forever:
  Wait until $\mathcal{C}_{TC}$ records a request $(\mathsf{id}, \mathsf{params}, \_)$:
  Fork a process of **Handle**$(\mathsf{id}, \mathsf{params})$
End

Figure 7: The Town Crier Relay $\mathcal{R}$.

# 5 Enhanced Protocol

The simplified payment-free protocol in Section **??** guarantees authenticity of data for an honest user. In Ethereum, however, computation is not free: Recall from that Ethereum employs *gas* to fuel contracts. This means that TC needs enough gas to deliver datagrams, so users must pay a fee to reimburse costs. If fees are not properly handled, a malicious relay (or malicious user) could prevent delivery of a datagram and/or cost an honest requester money for no gain. We discuss Section **??** how we address this issue throughout TC's design We also briefly discuss other enhancements to the basic TC protocol: Private and custom datagrams (Section **??**) and use of replication / voting to achieving robustness against SGX host or data source compromise (Section **??**).

## 5.1 Handling Fees in Ethereum

To address the above concern of attacks on TC fee management, we employ and prove the security of a novel two-currency resource-management system. This system causes requesters to make gas payments up front as ether. It converts ether to gas so as to prevent a malicious requester from exhausting TC's resources or a malicious TC from stealing an honest user's money. We now give some preliminaries and then explain our system.

**Execution model and notation.** We take Ethereum's gas model as described in Section **??**. We use the notation $\$g to denote gas and $\$f to denote non-gas currency. In both cases \$ is a type annotation and the letter denotes the numerical amount. For simplicity, our notation assumes that gas and normal currency adopt the same units (allowing us to perform arithmetic without conversion). We use the following identifiers to denote currency and gas amounts.

| | |
|---|---|
| $\$f$ | Currency a requester deposits to refund Town Crier's gas expenditure to deliver a datagram |
| $\$g_{\text{req}}$ $\$g_{\text{dvr}}$ $\$g_{\text{cncl}}$ | `GASLIMIT` when invoking **Request**, **Deliver**, or **Cancel**, respectively |
| $\$g_{\text{clbk}}$ | `GASLIMIT` for callback while executing **Deliver**, set to the max value that can be reimbursed |
| $\$G_{\text{min}}$ | Gas required for **Deliver** excluding callback |
| $\$G_{\text{max}}$ | Maximum gas TC can spent to invoke **Deliver** |
| $\$G_{\text{cncl}}$ | Gas needed to invoke **Cancel** |
| $\$G_{\emptyset}$ | Gas needed for **Deliver** on a cancelled request |

Note that $\$G_{\text{min}}$, $\$G_{\text{max}}$, $\$G_{\text{cncl}}$, and $\$G_{\emptyset}$ are system constants, $\$f$ is chosen by the requester (and may be malicious if the requester is dishonest), and $\$g_{\text{dvr}}$ is chosen by the TC Enclave when calling **Deliver**. Though $\$g_{\text{req}}$ and $\$g_{\text{cncl}}$ are set by the requester, we need not worry about the values. If they are too small, Ethereum will abort the transaction and there will not be a request or cancellation.

**Adversarial cases.** During the creation and fulfillment of any request, there are two untrusted parties: the requesting contract / user $\mathcal{C}_U$ and the TC Relay. In the TC implementation, we thus consider three cases:

- *Honest requester and Relay.* The requester must receive a valid authenticated response from TC.
- *Malicious requester and honest Relay.* TC must still be able to respond to requests from other (honest) users. Thus we must prevent a malicious user from interfering directly with other requests (which the payment-free protocol already does) or exhausting the financial resources of TC.

- *Honest requester and malicious Relay.* The requester cannot receive invalid data (which is also assured by the payment-free protocol) and the requester not have to pay computation that is not executed.

We formalize these properties in Section **??** and prove that our protocol provides these guarantees. We intentionally ignore the case where both the requester and the Relay are dishonest. If the requester is dishonest we need to not protect their request, and if the Relay is dishonest we cannot protect the TC system.

**Town Crier blockchain contract $\mathcal{C}_{TC}$ with fees**

**Request:** On recv $(\mathsf{params}, \mathsf{callback}, \$\mathtt{f}, \$g_{\text{req}})$ from some $\mathcal{C}_U$:
Assert $\$G_{\min} \leq \$\mathtt{f} \leq \$G_{\max}$
$\mathsf{id} := \text{Counter}$; $\text{Counter} := \text{Counter} + 1$
Store $(\mathsf{id}, \mathsf{params}, \mathsf{callback}, \$\mathtt{f}, \mathcal{C}_U)$
*// $\$\mathtt{f}$ held by contract*

**Deliver:** On recv $(\mathsf{id}, \mathsf{params}, \mathsf{data}, \$g_{\text{dvr}})$ from $\mathcal{W}_{TC}$:
(1) If isCanceled[id] and not isDelivered[id]
Set isDelivered[id]
(2) Send $\$G_{\emptyset}$ to $\mathcal{W}_{TC}$
Return
Retrieve stored $(\mathsf{id}, \mathsf{params}', \mathsf{callback}, \$\mathtt{f}, \_)$
*// abort if not found*
Assert $\mathsf{params} = \mathsf{params}'$ and $\$\mathtt{f} \leq \$g_{\text{dvr}}$
Set isDelievered[id]
(3) Send $\$\mathtt{f}$ to $\mathcal{W}_{TC}$
Set $\$g_{\text{clbk}} := \$\mathtt{f} - \$G_{\min}$
(4) Call callback(data) with $\$g_{\text{clbk}}$ max gas

**Cancel:** On recv $(\mathsf{id}, \$g_{\text{cncl}})$ from $\mathcal{C}_U$:
Retrieve stored $(\mathsf{id}, \_, \_, \$\mathtt{f}, \mathcal{C}'_U)$
*// abort if not found*
Assert $\mathcal{C}_U = \mathcal{C}'_U$
and $\$\mathtt{f} \geq \$G_{\emptyset}$
and isDelivered[id] not set
and isCanceled[id] not set
Set isCanceled[id]
(5) Send $(\$\mathtt{f} - \$G_{\emptyset})$ to $\mathcal{C}_U$ *// hold* $\$G_{\emptyset}$

Figure 8: Town Crier contract $\mathcal{C}_{TC}$ reflecting fees. The last argument of each entry point is the GASLIMIT provided. An honest requester sets $\$\mathtt{f}$ to be the gas required to execute **Deliver** including callback. Town Crier sets $\$g_{\text{dvr}} := \$G_{\max}$, but lowers the gas limit for callback ensure that no more than $\$\mathtt{f}$ is spent.

**Town Crier protocol with fees.** Our basic Town Crier system implements a policy where the requester pays for all gas needed and Town Crier effectively pays nothing. We now describe how this can be realized by modifying the payment-free protocol described in Section **??**.

**Program for Town Crier Enclave** $(\text{prog}_{\text{encl}})$

**Initialize**: [Same as Figure **??**]

**Resume:** On recv $(\mathtt{resume}, (\mathsf{id}, \mathsf{params}))$
[Same as Figure **??** except the last two lines:]
$\sigma := \Sigma.\text{Sign}(\text{sk}_{TC}, (\mathsf{id}, \mathsf{params}, \mathsf{data}, \$G_{\max}))$
Output $((\mathsf{id}, \mathsf{params}, \mathsf{data}, \$G_{\max}), \sigma)$

Figure 9: The Town Crier Enclave $\text{prog}_{\text{encl}}$.

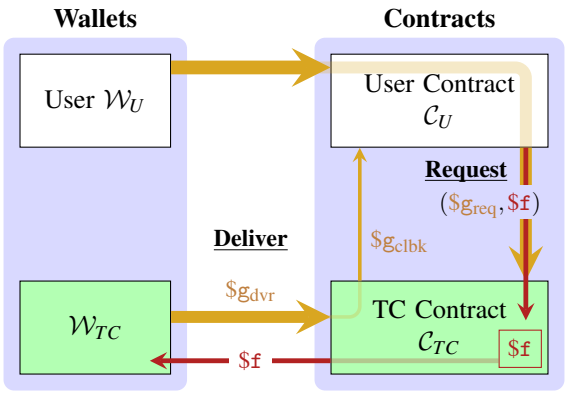


Figure 10: **Money Flow for a Delivered Request.** Red arrows denote flow of money and brown arrows denote gas limits for functions. The thickness of the line indicates the quantity of the resource. The $\$g_{\text{clbk}}$ arrow is thin because the value is limited to $\$\mathtt{f} - \$G_{\min}$.

- *Initialization.* We assume that Town Crier deposits at least $\$G_{\max}$ into the wallet $\mathcal{W}_{TC}$.
- *Town Crier blockchain contract.* Figure **??** describes the Town Crier blockchain contract reflecting fees. Since $\mathcal{W}_{TC}$ must invoke **Deliver**, TC will pays the gas cost. It sets the GASLIMIT $\$g_{\text{dvr}} := \$G_{\max}$. To ensure that the gas spent will not exceed the reimbursement available ($\$\mathtt{f}$), $\mathcal{C}_{TC}$ sets the GASLIMIT $\$g_{\text{clbk}}$ for the sub-call to callback to the reimbursement remaining: $\$\mathtt{f} - \$G_{\min}$.
- *Town Crier Relay.* The relay behavior does not change with the presence of fees. It still monitors the blockchain and whenever the contract $\mathcal{C}_{TC}$ stores a new request $(\mathsf{id}, \mathsf{params}, \_, \_, \_)$, it invokes $\text{prog}_{\text{encl}}$ with $\mathtt{resume}(\mathsf{id}, \mathsf{params})$.
- *Town Crier enclave.* We make the following small modification to the fee-free protocol. Instead of signing the tuple $(\mathsf{id}, \mathsf{params}, \mathsf{data})$ at the end of its execution, the enclave now signs the tuple $(\mathsf{id}, \mathsf{params}, \mathsf{data}, \$g_{\text{dvr}})$ where $\$g_{\text{dvr}} = \$G_{\max}$.
- *Requester.* An honest requester follows the protocol in Fig. **??** to verify the SGX attestation. Then she then prepares params and callback, sets $\$g_{\text{req}}$ to the

gas cost of **Request** with params, and $f to $G_{min} plus the cost of executing callback. The she invokes **Request**(params, callback, $f) and GASLIMIT $g_{req}.

If callback is not executed, the requester will invoke **Cancel** with argument id and gas limit $G_{cncl} to receive a partial refund. The honest requester will never invoke **Cancel** more than once for the same request and will never invoke **Cancel** with an id that corresponds to a different user's request.

### 5.2 Private and custom datagrams

In addition to ordinary datagrams, TC supports *private datagrams*, which involve requests where params includes ciphertexts under $pk_{TC}$. Private datagrams can thus enable confidentiality-preserving applications despite the public readability of the blockchain. *Custom datagrams*, also supported by TC, allow a contract to specify a particular web-scraping target, potentially involving multiple interactions, and thus greatly expand the range of possible relying contracts for TC. We give examples of both datagram types in Section **??**.

### 5.3 Enhanced robustness via replication

Our basic security model for TC assumes the ideal isolation model for SGX described above, as well as client trust in data sources. We nonetheless include hedges in TC, given various concerns about SGX security [**?**, **?**], and the possible fallibility of data sources. Briefly... [Ari: For Kyle to fill in]

## 6 TC Implementation Details

We now present further, system-level details on the TC contract $\mathcal{C}_{TC}$ and the two parts of the TC server, the Enclave and Relay.

### 6.1 TC Contract

We implement $\mathcal{C}_{TC}$ with fees as described in Section **??** in Solidity, a high-level language with JavaScript-like syntax which compiles to Ethereum Virtual Machine bytecode—the language Ethereum contracts use.

In order to handle the most general type of requests—including encrypted parameters—the $\mathcal{C}_{TC}$ implementation requires two parameter fields: a single byte specifying the type of request (e.g. flight status) and a byte array of user-specified size. This byte array is parsed and interpreted inside the Enclave, but is treated as an opaque byte array by $\mathcal{C}_{TC}$. For convenience, we include the timestamp of the current block as an implicit parameter.

To guard against the Relay tampering with request parameters, the $\mathcal{C}_{TC}$ protocol includes params as an argument to **Deliver** which validates against stored values. To reduce this cost for large arrays, we store and verify SHA3-256(typeByte||timestamp||paramArray). The Relay scrapes the raw values for the Enclave which computes the hash and includes it as an argument to **Deliver**.

### 6.2 TC Server

Using the recently released Intel SGX SDK [**?**], we implemented the TC Server as an SGX-enabled application in C++. In the programming model supported by the SGX SDK, the body of an SGX-enabled application runs as an ordinary user-space application, while a relatively small piece of security-sensitive code runs in the isolated environment of the SGX enclave.

The enclave portion of an SGX-enabled application may be viewed as a shared library exposing an API in the form of *ecalls* to be invoked by the untrusted application. Invocation of an ecall transfers control to the enclave; the enclave code runs until it either terminates and explicitly releases control, or some special event (e.g., exception) happens [**?**]. Again, as we assume SGX provides ideal isolation, the untrusted application cannot observe or alter the execution of ecalls.

Enclave programs can make *ocalls* to invoke functions defined outside of the enclave. An ocall triggers an exit from the enclave; control is returned once the ocall completes. As ocalls execute outside the enclave, they must be treated by enclave code as untrusted.

For TC, we recall that Fig. **??** shows the Enclave code $prog_{encl}$. Fig. **??** specifies the operation of the Relay, the untrusted code in TC, which we emphasize again provides essentially only network functionality. We now give details on the services in the Enclave and the Relay and describe their interaction, as summarized in Fig. **??**.

**The Enclave.** There are three components to the enclave code $prog_{encl}$: An HTTPS service, Web Scrapers, which interact with data sources, and a Request Handler, which services datagram requests.

*HTTPS Service.* We recall that the enclave does not have direct access to host network functionality. TC thus partitions HTTPS into a trusted layer, consisting of HTTP and TLS code, and an untrusted layer that provides low-layer network service, specifically TCP. This arrangement allows the enclave to establish a secure channel with a web server: The enclave itself performs the TLS handshake with a target server and performs all cryptographic operations internally, while the untrusted process acts as a network interface only. We ported a TLS library (mbedTLS) into the SGX environment, as well as HTTP code, which we minimized to meet the web-scraping re-

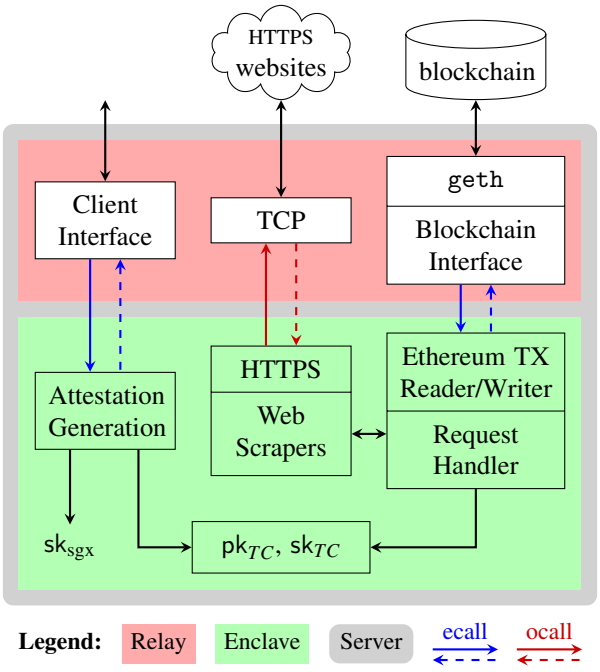


Figure 11: Components of TC Server

quirements of TC while keeping the TCB small. To verify certificates presented by remote servers, we hard-coded a collection of root CA certificates into the enclave code; in the first version of TC, the root CAs are identical to those in Chrome []. By using its internal, trusted wall-clock time, it is possible to verify that a certificate has not expired. (We briefly discuss revocation in Appendix **??**.)

*Web Scrapers.* We implemented scrapers for our examples in Section **??** in an ad hoc manner for our initial implementation of TC, and defer more principled, robust approaches to future work.

*Request Handler.* The Request Handler has two jobs: (1) Ingesting a datagram request by parsing it in the serialization format specified by Ethereum, decrypting it (if it's a private-datagram request), and dispatching its parameters to the right scraper; and (2) Returning the response to the request by generating an Ethereum transaction containing the requested datagram (and parameters), serializing it as a blockchain transaction, and signing it using $\mathsf{sk}_{TC}$. We implemented the Ethereum ABI and RLP, which respectively specify the serialization of arguments and transactions in Ethereum.

*Attestation Generation.* Recall in Section **??** we mentioned that an *attestation* is an *report* digitally signed by the Intel-provided Quoting Enclave (QE). Therefore two phases are involved in generating att. First, the Enclave calls `sgx_create_report` to generate a report with QE as the target enclave. Then the Relay forwards the report to QE and calls `sgx_get_quote` to get a signed version of the report, namely an attestation.

**The Relay.** The Relay encompasses three components: A Client Interface, which serves attestations and timestamps, OS services, including networking and time services, and a Blockchain Interface.

*Client Interface.* As described in Section **??**, a client starts using TC by requesting and verifying an attestation att and checking the correctness of the clock in the TC enclave using a fresh timestamp. The Client Interface caches att upon initialization of $\mathsf{prog}_{\text{encl}}$. When it receives a web request from a client for an attestation, it issues an ecall to the enclave to obtain a Unix timestamp signed using $\mathsf{sk}_{TC}$, which it returns to the client along with att. The client verify att using the Intel Attestation Service (IAS) [] and then verify the timestamp using $\mathsf{pk}_{TC}$ and check it using any trustworthy time service.

*OS services.* The Enclave relies on the Relay to access networking and wall-clock time provided by the OS and implemented as ocalls.

*Blockchain Interface.* The Relay's Blockchain Interface monitors the blockchain for incoming requests and places transactions on the blockchain in order to deliver datagrams. The Blockchain Interface incorporates an official Ethereum client, Geth [**?**]. This Geth client can be configured with a JSON RPC server. The Relay communicates with the blockchain indirectly via RPC calls to this server. For example, to insert a signed transaction, the Relay simply calls `eth_sendRawTransaction` with the byte array of the serialized transaction. We emphasize that as the enclave holds $\mathsf{sk}_{TC}$, transactions are signed within the enclave.

## 7 Security Analysis

Proofs of theorems in this section appear in Appendix **??**.

**Authenticity.** Intuitively, authenticity means that an adversary (including a corrupt user, relay, or collusion thereof) cannot convince $\mathcal{C}_{TC}$ to accept a datagram that differs from the expected content obtained by crawling the specified url at the specified time. In our formal definition, we assume that the user and $\mathcal{C}_{TC}$ behave honestly. Recall that the user must verify upfront the attestation $\sigma_{\text{att}}$ that vouches for the enclave's public key $\mathsf{pk}_{\text{TC}}$.

**Definition 1** (Authenticity of Data Feed)**.** *We say that the TC protocol satisfies* authenticity of data feed *if, for any polynomial-time adversary that can interact arbitrarily with $\mathcal{F}_{sgx}$, it cannot persuade an honst verifier to accept* $(\mathsf{pk}_{TC}, \sigma_{att}, \mathsf{params} := (\mathsf{url}, \mathsf{pk}_{url}, T), \mathsf{data}, \sigma)$ *where* data *is not the contents of* url *with the public key* $\mathsf{pk}_{url}$ *at time T. More formally, for any probabilistic polynomial-time*

*adversary $\mathcal{A}$,*

$$\Pr\left[\begin{array}{l}(\mathsf{pk}_{TC}, \sigma_{att}, \mathsf{id}, \mathsf{params}, \mathsf{data}, \sigma) \leftarrow \mathcal{A}^{\mathcal{F}_{sgx}}(1^{\lambda}) : \\ \quad \big(\Sigma_{sgx}.\mathsf{Verify}(\mathsf{pk}_{\mathrm{sgx}}, \sigma_{att}, (\mathsf{prog}_{encl}, \mathsf{pk}_{TC})) = 1\big) \wedge \\ \quad (\Sigma.\mathsf{Verify}(\mathsf{pk}_{TC}, \mathsf{id}, \mathsf{params}, \mathsf{data}) = 1) \wedge \\ \quad \mathsf{data} \neq \mathsf{prog}_{encl}(\mathsf{params})\end{array}\right] \leq \mathsf{negl}(\lambda)$$

**Theorem 1** (Authenticity)**.** *Assume that $\Sigma_{sgx}$ and $\Sigma$ are secure signature schemes (recall that we follow Shi et al. [elaine: cite] who show how to abstractly model SGX's group signature as a regular signature scheme under a manufacturer public key $\mathsf{pk}_{\mathrm{sgx}}$), then, the above protocol achieves authenticity of data feed by Definition* **??**.

**Fee Safety.** Our protocol in Section **??** ensures that an honest Town Crier will not run out of money and that an honest requester will not pay excessive fees.

**Theorem 2** (Gas neutrality for Town Crier)**.** *If the TC Relay is honest, Town Crier's wallet $\mathcal{W}_{TC}$ will have at least $\$\mathtt{G}_{max}$ remaining after each* **Deliver** *call.*

An honest user should only have to pay for computation that is executed honestly on their behalf. If a valid datagram is delivered, this is a constant value plus the cost of executing callback. Otherwise the request should be able to recover the cost of executing **Deliver**. For Theorem **??** to hold, $\mathcal{C}_{TC}$ must retain a small fee on cancellation, but we allow the user to recover all but a small constant amount. We formalize this intuition in Theorem **??**

**Theorem 3** (Fair Expenditure for Honest Requester)**.** *For any* params *and* callback*, let $\$\mathtt{G}_{req}$ and $\$\mathtt{F}$ be the respective values chosen by an honest requester for $\$\mathtt{g}_{req}$ and $\$\mathtt{f}$ when submitting the request* $(\mathsf{params}, \mathsf{callback}, \$\mathtt{f}, \$\mathtt{g}_{req})$*. For any such request submitted by an honest user $\mathcal{C}_U$, one of the following holds:*

- callback *is invoked with a valid datagram matching the request parameters* params *and the requester spends at most $\$\mathtt{G}_{req} + \$\mathtt{G}_{cncl} + \$\mathtt{F}$.*
- *The requester spends at most $\$\mathtt{G}_{req} + \$\mathtt{G}_{cncl} + \$\mathtt{G}_{\emptyset}$.*

**Other security concerns.** We address concerns about attacks against the basic SGX isolation model embraced in the basic TC protocol in Section **??**. A threat we do not address in TC is the risk of traffic analysis by a network adversary or compromised Relay against confidential applications (e.g., with private datagrams), but briefly discuss the issue in Section **??**.

# 8 Applications: Requesting Contracts

We built and implemented several showcase applications which we plan to launch together with Town Crier. We give a description of these applications in this section, and show experimental results in Section **??**. We refer the reader to Appendix **??** for more details on these applicatins, as well as $\mathcal{C}_U$ code samples that demontrate first-hand experience of how to use Town Crier's service.

**Financial derivative** (CashSettledPut)**.** Financial derivatives are among the most commonly cited smart contract applications, and exemplify the need for a data feed on financial instruments. We implemented an example contract CashSettledPut for a *cash-settled put option*. This is an agreement for one party to buy an asset from the other at an agreed upon price on or before a particular date. It is "cash-settled" in that the sale is implicit, i.e. no asset changes hands, only cash reflecting the asset's value.

**Flight insurance** (FlightIns)**.** Flight insurance indemnifies a purchaser should her flight be delayed or canceled. We have implemented a simple flight insurance contract called FlightIns. Our implementation showcases TC's *private-datagram* feature to address an obvious concern: customers may not wish to reveal their travel plans publicly on the blockchain. Roughly speaking, a customer submits to $\mathcal{C}_{TC}$ a request $\mathsf{Enc}_{\mathsf{pk}_{TC}}(\mathsf{req})$ encrypted under Town Crier enclave's public key $\mathsf{pk}_{TC}$. The enclave decrypts the request and checks its well-formedness (e.g., the request is submitted sufficiently long before the flight time). If well-formed, the enclave will fetch the flight information website at a specified later time, and send to $\mathcal{C}_{TC}$ a datagram indicating whether the flight is cancelled. Finally, to avoid leaking information through timing (e.g., when the flight information website is accessed), random delays can be introduced to mitigate the information leakage.

**Steam Marketplace** (SteamTrade)**.** Authenticated data feeds and smart contracts can enable fair exchange of digital goods between Internet users who do not have pre-established trust. To do this, we have developed an example application supporting fair trade of virtual items for the Steam [**?**], an onling gaming platform that supports thousands of games and maintains its own marketplace, where users can trade, buy, and sell games and other virtual items. We implement a contract for the sale of games and items for ether that showcases TC's support for *custom datagrams* through the use of Steam's access-controlled API. In our implementation, the buyer sends $\mathsf{Enc}_{\mathsf{pk}_{TC}}(\text{account credentials}, \mathsf{req})$ to $\mathcal{C}_{TC}$, such that Town Crier's enclave can log in as the buyer and determine from the web-page whether the virtual item has been shipped.

## 9 Experiments

We evaluated the performance of TC on a Dell Inspiron 13-7359 laptop equipped with an Intel i7-6500U CPU and 8.00GB memory. This model is chosen because it is one of the few SGX-enabled systems available on the market at the writing of this paper. Although the hosting platform is just moderately preferment, we show that on a single host, the current implementation is able to handle transactions generated by Ethereum blockchain in real time. [elaine: todo: modify the previous sentence after Fan's new experiments are done.]

### 9.1 Enclave Response Time

We measured the enclave response time for handling a request. Here "enclave response time" is defined as the difference between 1) the time the relay decides to send a request to an SGX enclave; and 2) the time the relay gets a response back from the SGX enclave. For each application, we repeat the experiment 5 times, and report the mean and the standard deviation. [elaine: change accordingly after Fan puts in the std dev]

Table **??** summarizes the total enclave response time as well as its breakdown. The table suggests that for the three applications we implemented the enclave response time ranges from **192 ms** to **1309 ms**. [elaine: update numbers after Fan updates table] The response time is clearly dominated by the web scraper time, i.e., the time it takes to fetch the requested information from a website. Among the three applications evaluated, SteamTrade has the longest web scraper time, since in the case of SteamTrade, the web scraper must interact with the website over multiple roundtrips to fetch the desired datagram.

### 9.2 Transaction Throughput

We performed a sequence of experiments running 1 to 20 enclaves on a single SGX-enable laptop machine, and investigated the transaction throughput enabled by a single SGX machine. Note that SGX allows at most 20 enclaves on the specific machine model we used. Figure [elaine: refer to Fan's new figure] shows that for the three applications evaluated, a single SGX machine can handle [elaine: blah to blah] tx/sec.

We give several meaningful data points of comparison to show why a single SGX machine suffices to handle the load of today's blockchains: Ethereum as of today handles [elaine: fill in] tx/sec on average. Bitcoin today handles [elaine: fill in] tx/sec, and its maximum throughput (when the full block size is utilized) would be roughly 7 tx/sec. We know of no measurement study that investigates the fundamental throughput bound of the Ethereum peer-to-peer network. However, for Bitcoin, it has been shown that without redesigning the protocol, Bitcoin cannot scale beyond 42 tx/sec [elaine: double check this number] by simple reparametrization of its block size. [elaine: cite] Therefore, Town Crier will not be a throughput bottleneck when used with decentralized blockchains — it appears more imminent to scale up the blockchain protocols themselves.

Last but not the least, our experiments indicate that *there is no noticeable slowdown in terms of enclave response time when multiple enclaves are run* on a single machine. [elaine: Fan: double check]

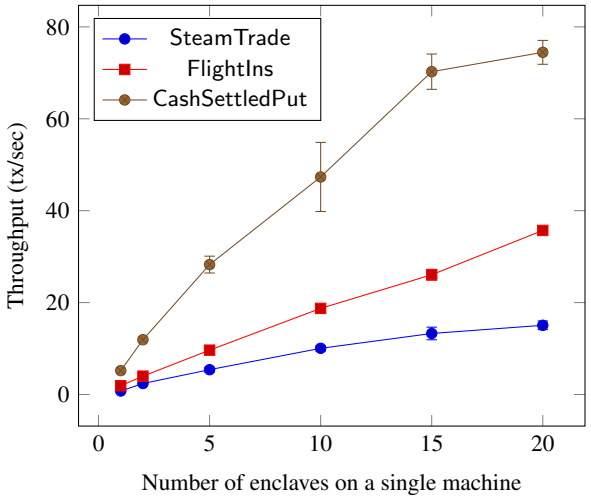


Figure 12: Throughput is measured on a single SGX-enabled machine. The x-axis is the number of enclaves and the y-axis is the number of transactions processed per second. We ran five rounds of experiments and error bars show the standard deviation.

### 9.3 Gas Costs

Currently 1 gas costs $5/10^8$ ether, so at the exchange rate of $5 for 1 ether, $1 buys 4 million gas. Here we provide costs for the components of our implementation.

The callback-independent portion of **Deliver** costs about 35,000 gas (0.875¢), so this is the value of $\$G_{min}$. We set $\$G_{max} = 3{,}100{,}000$ gas (77.5¢) as this is approximately Ethereum's maximum `GASLIMIT`. The cost for executing **Request** is approximately 120,000 gas (3¢) of fixed cost, plus 2500 gas (0.06¢) for every 32 bytes of request parameters. The cost for executing **Cancel** is [elaine: fill in] gas ([Ethan: fill]¢) including the gas cost $\$G_{cncl}$ and the refund $\$G_{\emptyset}$ paid to Town Crier should **Deliver** be called after **Cancel**.

| | CashSettledPut | | FlightIns | | SteamTrade | |
|---|---|---|---|---|---|---|
| | time(ms) | proportion | time(ms) | proportion | time(ms) | proportion |
| Context switch | 0.41 | 0.21% | 0.44 | 0.08% | 0.89 | 0.07% |
| Web scraper | 181.03 | 94.12% | 512.32 | 98.02% | 1297.73 | 99.11% |
| Sign | 10.63 | 5.53% | 9.64 | 1.85% | 10.45 | 0.80% |
| Serialization | 0.28 | 0.14% | 0.25 | 0.05% | 0.32 | 0.02% |
| Total | 192.35 | 100.00% | 522.65 | 100.00% | 1309.39 | 100.00% |

Table 1: Running time of handling a request [elaine: Fan to 1) add std dev; 2)update context switch time to include time switching out too]

| | CashSettledPut | FlightIns | SteamTrade |
|---|---|---|---|
| Delivery before issuing cancel | ?? | ?? | ?? |
| Cancel arrived after delivery | | | |
| Cancel, no delivery | | | |

Table 2: **Callback-independent** portion of the gas expenditure, translated to USD. Here the difference between the gas expenditure across applications is mainly caused by the difference in length of input parameters and output datagrams. [elaine: ethan: check]

### 9.4 Resilience against Component Compromise

For the FlightIns application, we implemented and evaluated two modes of majority voting:

- 2-out-of-3 majority voting within the enclave which provides robustness against data source compromise. Specifically, the enclave scraped three different data sources in three parallel threads, namely, [elaine: fill in]. Our experiments suggest that the enclave response time is [elaine: fill] ms in this case (*c.f.* [elaine: fill] ms, [elaine: fill] ms, and [elaine: fill] ms for each single data source). There is no change in the gas cost in this case.
- 2-out-of-3 majority voting within the requester contract, which provides robustness against SGX compromise. To support this, we ran three instances of SGX enclaves all scraping the same data source. The main change in this scenario is that the gas cost would increase from [elaine: fill] cents to [elaine: fill] cents. In particular, the increase is just slightly more than 3x because [elaine: Ethan to fill in a very short explanation].

### 9.5 Offline Measurements

Recall that a one-time setup operation is required for the enclave to be established, and an attestation generated. Our measurement suggests that the enclave establishment takes [elaine: fill] ms, and the attestation generation takes 83.4 ms (among which 18.5 ms is spent on report generation, and 64.9 ms on quote generation).

Recall also that since SGX only reports a relative clock with respect to some reference point, a user must perform offline clock calibration to translate the relative clock values into absolute ones. The clock precision is decided by the end-to-end latency for performing this clock calibration. Our experiments show that the latency is [elaine: fill] ms between when a relay sends a clock calibration request to the enclave, and when it obtains a response. The majority of this [elaine: fill] ms is spent on [elaine: blah operation] which took [elaine: blah] ms. In practice, the clock precision is also affected by the wide-area network latency which is typically in the [elaine: blah] range.

## 10 Related Work

Several data feeds are deployed today for smart contract systems such as Ethereum. Examples include PriceFeed [**?**] and Oraclize.it [**?**]. The latter achieves distributed trust by using a second service called TLSnotary [**?**], which digitally signs webpages. These services, however, ultimately rely on the reputations of their (small) providers to ensure data authenticity. To address this problem, systems such as SchellingCoin [**?**] and Augur [**?**] rely on mechanisms such as prediction markets to decentralize trust. Prediction markets often rely on human input, though, constraining their scope. They have not yet seen widespread use in cryptocurrencies.

Despite an active developer community, research community exploration of smart contracts has been very limited to date. Work includes off-block contract execu-

tion for confidentiality [**?**], and, more tangentially, exploration of e.g., randomness sources in [**?**]. The only exploration relating to data feeds of which we're aware takes the form of consuming (criminal) applications in [].

SGX is similarly in its infancy. While a Windows SDK [**?**] and programming manual [**?**] have just been released, a number of pre-release papers have already explored SGX, e.g., [**?**, **?**, **?**, **?**, **?**, **?**]. Researchers have demonstrated applications include enclave execution of legacy (non-SGX) code [**?**] and use of SGX in a distributed setting for map-reduce computations [**?**]. Several works have exposed shortcomings of the security model for SGX [**?**, **?**, **?**], including side-channel attacks and other attacks against enclave state.

## 11 Conclusion

## A Showcase Applications and Code Samples

We now elaborate on the applications described in Section **??** and we show a short Solidity code sample for one of these applications, to demonstrate first-hand what a requester contract would look like to call Town Crier's authenticated data feed service. [elaine: TODO: actually put in syntax highlighted code example here.]

**Financial derivative** (CashSettledPut)**.** Financial derivatives are among the most commonly cited smart contract applications, and exemplify the need for a data feed on financial instruments. We implemented an example contract CashSettledPut for a *cash-settled put option*. This is an agreement for one party to buy an asset from the other at an agreed upon price on or before a particular date. It is "cash-settled" in that the sale is implicit, i.e. no asset changes hands, only cash reflecting the asset's value. In our implementation, the issuer of the option specifies a strike price $P_S$, expiration date, unit price $P_U$, and maximum number of units $M$ she is willing to sell. A customer may send a request to the contract specifying the number $X$ of option units to be purchased and containing the associated fee $(X \cdot P_U)$. A customer may then exercise the option by sending another request prior to the expiration date. CashSettledPut calls TC to retrieve the closing price $P_C$ of the underlying instrument on the day the option was exercised, and pays the customer $X \cdot (P_S - P_C)$. To ensure sufficient funding to pay out, the contract must be endowed with ether value at least $M \cdot P_S$.

**Flight insurance** (FlightIns)**.** Flight insurance indemnifies a purchaser should her flight be delayed or canceled. We have implemented a simple flight insurance contract called FlightIns. Our implementation showcases TC's *private-datagram* feature to address an obvious concern: customers may not wish to reveal their travel plans publicly on the blockchain.

An insurer stands up FlightIns with a specified policy fee, payout, and lead time $\Delta T$. ($\Delta T$ is set large enough to ensure that a customer can't anticipate flight cancellation or delay due to weather, etc.) To purchase a policy, a customer sends the FlightIns a ciphertext $C$ under the TC's pubic key $\mathsf{pk}_{TC}$ of the ICAO flight number $FN$ and scheduled time of departure $T_D$ for her flight, along with the policy fee. FlightIns sends TC a private-datagram request containing the current time $T$ and the ciphertext $C$. TC decrypts $C$ and checks that the lead time meets the policy requirement, i.e., that $T_D - T \geq \Delta T$. TC then scrapes a flight information data source several hours after $T_D$ to check the flight status, and returns to FlightIns predicates on whether the lead time was valid and whether the flight has been delayed or cancelled. If both

predicates are true, then FlightIns returns the payout to the customer. Note that $FN$ is never exposed in the clear.

Despite the use of private datagrams, FlightIns as described here still poses a privacy risk, as the *timing* of the predicate delivery by TC leaks information about $T_D$, which may be sensitive information; this, and the fact that the payout is publicly visible, could also indirectly reveal $FN$. FlightIns addresses this issue by including in the private datagram request another parameter $t > T_D$ specifying the time at which predicates should be returned. By randomizing $t$ and making $t - T_D$ sufficiently large, FlightIns can substantially reduce the leakage of timing information.

**Steam Marketplace (SteamTrade).** Steam [**?**] is an online gaming platform that supports thousands of games and maintains its own marketplace, where users can trade, buy, and sell games and other virtual items. We implement a contract for the sale of games and items for ether that showcases TC's support for custom datagrams through the use of Steam's access-controlled API.

A user intending to sell items creates a contract SteamTrade with his Steam account number $ID_S$, a list $L$ of items for sale and a price in ether for each. In order to purchase some subset $L_B$ of the items, a buyer first uses a Steam client to create a trade offer requesting each item $i \in L_B$. The buyer then submits to SteamTrade a ciphertext $C$ under the TC's public key $\mathsf{pk}_{TC}$ of his Steam API key $K_S$, his account number $ID_B$, the list of items $L_B$, and a time period $T_B$ indicating how long the seller has to respond to the offer, along with an amount of ether equivalent to the sum of the specified prices of all items in $L_B$. SteamTrade sends TC a custom datagram request containing the current time $T$, the account number $ID_S$, and the ciphertext $C$. TC decrypts $C$, delays for time $T_B$, then retrieves all trades between the two accounts identified by $ID_S$ and $ID_B$ within that time period by using the provided API key $K_S$. TC verifies whether or not a trade exactly matching the items in $L_B$ successfully occurred between the two accounts and returns the result to SteamTrade. If such a trade occurred, SteamTrade sends the buyer's ether to the seller's account. Otherwise the buyer's ether is refunded.

# B More Details on Formal Modeling

## B.1 SGX Formal Modeling

As mentioned earlier, we adopt the UC model of SGX proposed by Shi et al. [**?**] In particular, their abstraction captures a subset of the features of Intel SGX. The main idea behind the UC modeling by Shi et al. [**?**] is to think of SGX as a trusted third party defined by a global functionality $\mathcal{F}_{\text{sgx}}$ (see Figure **??** of Section **??**).

**Modeling choices.** For simplicity, the $\mathcal{F}_{\text{sgx}}$ model currently does not capture the issue of revocation. In this case, as Shi et al. point out, we can model SGX's group signature simply as a regular signature scheme $\Sigma_{\text{sgx}}$, whose public and secret keys are called "manufacturer keys" and denoted $\mathsf{pk}_{\text{sgx}}$ and $\mathsf{sk}_{\text{sgx}}$ (i.e., think of always signing with the 0-th key of the group signature scheme). We adopt this notational choice from Shi et al. [**?**] for simplicity. Later when we need to take revocation into account, it is always possible to replace this signature scheme with a group signature scheme in the modeling.

The $\mathcal{F}_{\text{sgx}}(\Sigma_{\text{sgx}})$ functionality described by Shi et al. [**?**] is a global functionality shared by all protocols, parametrized by a signature scheme $\Sigma_{\text{sgx}}$. This global $\mathcal{F}_{\text{sgx}}$ is meant to capture all SGX machines available in the world, and keeps track of multiple execution contexts for multiple enclave programs, happening on different SGX machines in the world. For convenience, this paper adopts a new notation $\mathcal{F}_{\text{sgx}}(\Sigma_{\text{sgx}})[\mathsf{prog}_{\text{encl}}, \mathcal{R}]$ to denote one specific execution context of the global $\mathcal{F}_{\text{sgx}}$ functionality where the enclave program in question is $\mathsf{prog}_{\text{encl}}$, and the specific SGX instance is attached to a physical machine $\mathcal{R}$. This specific context $\mathcal{F}_{\text{sgx}}(\Sigma_{\text{sgx}})[\mathsf{prog}_{\text{encl}}, \mathcal{R}]$ ignores all parties' inputs except those coming from $\mathcal{R}$. We often omit writing $(\Sigma_{\text{sgx}})$ without risk of ambiguity.

**Operations.** $\mathcal{F}_{\text{sgx}}$ captures the following features:

- *Initialize.* Initialization is run only once. Upon initialization, $\mathcal{F}_{\text{sgx}}$ runs the initialization part of the enclave program denoted $\mathsf{outp} := \mathsf{prog}_{\text{encl}}.\textbf{Initialize}()$. Then, $\mathcal{F}_{\text{sgx}}$ attests to the code of the enclave program $\mathsf{prog}_{\text{encl}}$ as well as $\mathsf{outp}$. The resulting attestation is denoted $\sigma_{\text{att}}$.

- *Resume.* When "resume" is invoked, $\mathcal{F}_{\text{sgx}}$ calls $\mathsf{prog}_{\text{encl}}.\textbf{Resume}$ on the input parameters denoted $\mathsf{params}$. $\mathcal{F}_{\text{sgx}}$ outputs whatever $\mathsf{prog}_{\text{encl}}.\textbf{Resume}$ outputs. $\mathcal{F}_{\text{sgx}}$ is stateful, i.e., allowed to carry state between "initialize" and multiple "resume" invocations.

Finally, we remark that this formal model by Shi et al. is speculative, since we know of no formal proof that Intel's SGX does securely realize this abstraction (or realize any useful formal abstraction at all for that matter) — in fact, available public documentations of SGX does not provide sufficient information for making such formal proofs. As such, the formal model by Shi et al. appears to be the best available tool for us to formally reason about security for SGX-based protocols. Shi et al. leave it as an open question to design secure processors with clear formal specifications, such that they can be used in the design of larger protocols/systems supporting formal reasoning of security. We refer the readers to

Shi et al. [**?**] for a more detailed description of the UC modeling of Intel SGX.

## B.2 Blockchain Formal Modeling

Our protocol notation adopts the formal blockchain framework recently proposed by Kosba et al. [**?**]. Besides UC modeling of blockchain-based protocols, Kosba et al. [**?**] also design a modular notational system that is intuitive and factors out tedious but common features inside functionality and protocol wrappers (e.g., modeling of time, pseudonyms, adversarial reordering of messages, a global ledger). The advantages of adopting Kosba et al.'s notational system is compelling: the blockchain contracts and user-side protocols in this paper are intuitive to understand on their own; moreover, they are endowed with precise, formal meaning when we apply the blockchain wrappers defined by Kosba et al.

**Technical subtleties.** While Kosba et al.'s formal blockchain model is applicable for the most part, we point out a subtle mismatch between the formal blockchain model of Kosba et al. [**?**] and the real-world instantiation of blockchains such as Ethereum (and Bitcoin for that matter). The design of Town Crier is secure in a slightly modified version of the blockchain model that more accurately reflects the real-world Ethereum instantiation of a blockchain.

Recall that one tricky detail for the gas-aware version of the Town Crier contract arises from the need to deal with with **Deliver** arriving after **Cancel**. In the formal blockchain model proposed by Kosba et al. [**?**], we can easily get away with this issue by introducing a timeout parameter $\mathsf{T}_{\text{timeout}}$ that the requester attaches to each datagram request. If the datagram fails to arrive before $\mathsf{T}_{\text{timeout}}$, the requester can call **Cancel** any time after $\mathsf{T}_{\text{timeout}} + \Delta T$. On the surface, this seems to ensure that no **Deliver** will be invoked after **Cancel** assuming Town Crier is honest.

However, we did not adopt this approach due to a technical subtlety that arises in this context, namely, the fact that the Ethereum blockchain does not perfectly match the formal blockchain model specified by Kosba et al [**?**]. Specifically, the blockchain model by Kosba et al. assumes that every message (i.e., transaction) will be delivered to the blockchain by the end of each epoch and that the adversary cannot drop any message. In practice, however, Ethereum adopts a dictatorship strategy in the mining protocol, and the winning miner for an epoch can censor transactions for this specific epoch, and thus effectively this transaction will be deferred to later epochs. Further, in case there are more incoming transactions than the block size capacity of Ethereum, a backlog of transactions will build up, and similarly in this case there is also no guarantee of ordering of backlogged transactions. Due to these considerations, we defensively design our Town Crier contract such that gas neutrality is attained for Town Crier, even if the **Deliver** transaction arrives after **Cancel**.

# C Proofs of Security Analysis

This section contains the proofs of the theorems we stated in Section **??**

*Authenticity (sketch).* We show that if the adversary $\mathcal{A}$ succeeds in a forgery with non-negligible probability, we can construct an adversary $\mathcal{B}$ that can either break $\Sigma_{\text{sgx}}$ or $\Sigma$ with non-neligible probability. We consider two cases. The reduction $\mathcal{B}$ will flip a random coin to guess which case it is, and if the guess is wrong, simply abort.

- Case 1: $\mathcal{A}$ outputs a signature $\sigma$ that uses the same $\mathsf{pk}_{\text{TC}}$ as the SGX functionality $\mathcal{F}_{\text{sgx}}$. In this case, $\mathcal{B}$ will try to break $\Sigma$. $\mathcal{B}$ interacts with a signature challenger $\mathsf{Ch}$ who generates some $(\mathsf{pk}^*, \mathsf{sk}^*)$ pair, and gives to $\mathcal{B}$ the public key $\mathsf{pk}^*$. $\mathcal{B}$ simulates $\mathcal{F}_{\text{sgx}}$ by implicitly letting $\mathsf{pk}_{\text{TC}} := \mathsf{pk}^*$. Whenever $\mathcal{F}_{\text{sgx}}$ needs to sign a query, $\mathcal{B}$ passes the signing query onto the signature challenger $\mathsf{Ch}$.

  Since $\mathsf{data} \neq \mathsf{prog}_{\text{encl}}(\mathsf{params})$, $\mathcal{B}$ cannot have queried $\mathsf{Ch}$ on a tuple of the form $(\_, \mathsf{params}, \mathsf{data})$. Therefore, $\mathcal{B}$ simply outputs what $\mathcal{A}$ outputs (suppressing unnecessary terms) as the signature forgery.

- Case 2: $\mathcal{A}$ outputs a signature $\sigma$ that uses a different $\mathsf{pk}_{\text{TC}}$ as the SGX functionality $\mathcal{F}_{\text{sgx}}$. In this case, $\mathcal{B}$ will seek to break $\Sigma_{\text{sgx}}$. $\mathcal{B}$ interacts with a signature challenger $\mathsf{Ch}$, who generates some $(\mathsf{pk}^*, \mathsf{sk}^*)$ pair, and gives to $\mathcal{B}$ the public key $\mathsf{pk}^*$. $\mathcal{B}$ simulates $\mathcal{F}_{\text{sgx}}$ by implicitly setting $\mathsf{pk}_{\text{sgx}} := \mathsf{pk}^*$. Whenever $\mathcal{F}_{\text{sgx}}$ needs to make a signature with $\mathsf{sk}_{\text{sgx}}$, $\mathcal{B}$ simply passes the signature query onto $\mathsf{Ch}$. In this case, in order for $\mathcal{A}$ to succeed, it must produce a valid signature $\sigma_{\text{att}}$ for a different public key $\mathsf{pk}'$. Therefore, $\mathcal{B}$ simply outputs this as a signature forgery.

□

*Proof of Gas neutrality for Town Crier (sketch).* By assumption, $\mathcal{W}_{TC}$ is seeded with at least $\$\mathbb{G}_{\max}$ money. Thus it suffices to prove that, given an honest Relay, $\mathcal{W}_{TC}$ will have at least as much money after invoking **Deliver** as it did before.

The enclave will never submit a response for the same $\mathsf{id}$ more than once. **Deliver** only responds to messages from $\mathcal{W}_{TC}$, and $\mathsf{isDelivered}[\mathsf{id}]$ is only set inside **Deliver**, so therefore we know that $\mathsf{isDelivered}[\mathsf{id}]$ is not set for this $\mathsf{id}$. We now consider the case where $\mathsf{isCanceled}[\mathsf{id}]$ is set upon invocation of **Deliver** and the case where it is not.

- isCanceled[id] *not set:* In this case the predicate on line (1) of the protocol returns `false`. Because the Relay is honest, id exists and params = params′. The enclave always provides $\$g_{dvr} = \$G_{max}$ (which it has by assumption) and **Request** ensures that $\$f \leq \$G_{max}$. Thus, coupled with the knowledge that isDeliver[id] is not set, all assertions pass and we progress through lines (3) and (4). Now we must show that at line (3) $\mathcal{C}_{TC}$ had $\$f to send and that the total gas spend to execute **Deliver** does not exceed $\$f.

  To see that $\mathcal{C}_{TC}$ had sufficient funds, we note that upon entry to **Deliver**, both isDelivered[id] and isCanceled[id] must have been unset. The first because the enclave will not attempt to deliver the same id more than once and isDeliver is only set in **Deliver** for the id being delivered. The second because, given the first, if isCanceled[id] were set, the predicate on line (1) would have returned true, thus sending execution on a path that would not encounter (4). This means that line (5) was never reached because the preceding line sets isCanceled[id]. Because (2), (3), and (5) are the only lines that remove money from $\mathcal{C}_{TC}$ and $\$f was deposited as part of **Request**, it must be the case that $\$f is still in the contract.

  To see how much gas is spent, we first note that $\$G_{min}$ is defined to be the amount of gas needed to execute **Deliver** along this execution path not including line (4). Since $\$g_{clbk}$ is set to $\$f - \$G_{min}$ and line (4) is limited to using $\$g_{clbk}$ gas, the total gas spent on this execution of **Deliver** is at most $\$G_{min} + (\$f - \$G_{min}) = \$f$.

- isCanceled[id] *is set:* Here the predicate on line (1) returns `true`. Along this execution path $\mathcal{C}_{TC}$ sends $\mathcal{W}_{TC}$ $\$G_{\emptyset}$ and quickly returns. $\$G_{\emptyset}$ is defined as the amount of gas necessary to execute this execution path, so we need only show that $\mathcal{C}_{TC}$ has $\$G_{\emptyset}$ available to send.

  Because isCanceled[id] is set, it must be the case that Cancel was invoked with id and reached line (5). Gas exhaustion in Cancel is not a concern because it would abort and revert the entire invocation. This is only possible if the data retrieval and all assertions in Cancel succeed. In particular, this means that id corresponds to a valid request which deposited $\$f. Line (5) returns $\$f - \$G_{\emptyset}$ to $\mathcal{C}_U$, but it leaves $\$G_{min}$ from the original $\$f. Moreover, if **Cancel** is invoked multiple times with the same id, all but the first will fail due to the assert that isCanceled[id] is not set and the fact that any invocation that reaches (5) will set isCanceled for that id. Since only lines (2), (3), and (5) can remove money from $\mathcal{C}_{TC}$ and line (3) will never be called in this case, there will still be exactly $\$G_{min}$ available when this invocation of **Deliver** reaches line (2).

□

*Proof of Fair Expenditure for Honest Requester (sketch).* $\mathcal{C}_U$ is honest, so she will spend $\$G_{req}$ to invoke **Request**(params, callback, $\$F). Ethereum does not allow money to change hands without the payer explicitly sending money. Therefore we must only examine the explicit function invocations and monetary transfers initiated by $\mathcal{C}_U$ in connection with the request. It is impossible for $\mathcal{C}_U$ to lose more money than she gives up in these transactions even if TC is malicious.

- *Request Delivered:* If protocol line (4) is reached, then we are guaranteed that params = params′ and $\$g_{dvr} \geq \$F$. By Theorem **??**, the datagram must therefore be authentic for params. Because $\$F is chosen honestly for callback, $\$F - \$G_{min}$ is enough gas to execute callback, so callback will be invoked with a datagram that is a valid and matches the request parameters.

  In this case, the honest requester will have spent $\$G_{req}$ to invoke **Request** and $\$F in paying TC's cost for **Deliver**. The requester may have also invoked **Cancel** at most once at the cost of $\$G_{cncl}$. While $\mathcal{C}_U$ may not receive any refund due to **Cancel** aborting, $\mathcal{C}_U$ will still have spent at most $\$G_{req} + \$G_{cncl} + \$F$.

- *Request not Delivered:* The request not being delivered means that line (4) is never reached. This can only happen if **Deliver** is never called with a valid response or if isCanceled[id] is set before deliver is called. The only way to set isCanceled[id] is for $\mathcal{C}_U$ to invoke **Cancel** with isDelivered[id] not set. If deliver is not executed, we assume that an honest requester will eventually invoke **Cancel**, so this case will always reach line (5). When line (5) is reached, then $\mathcal{C}_U$ will have spent $\$G_{req} + \$F$ while executing **Request**, and spent $\$G_{cncl}$ in **Cancel** and will attempt to retrieve $\$F - \$G_{\emptyset}$.

  The retrieval will succeed because $\mathcal{C}_{TC}$ will always have the funds to send $\mathcal{C}_U$ $\$F - \$G_{\emptyset}$. Since line (5) is reached, it must be the case the isDelivered[id] is not set. This means that neither lines (2) nor (3) were reached since the line before each sets isDelivered[id]. The lines preceding those two and (5) are the only lines that remove money from the contract. Line (5) cannot have been reached before because $\mathcal{C}_U$ is assumed to be honest, so she will not invoke **Cancel** twice for the same request and if any other user invokes **Cancel** for this request, the $\mathcal{C}_U = \mathcal{C}'_U$ assertion will fail and the invocation will abort before line (5). Because none of (2), (3), or (5) has been reached before and $\mathcal{C}_U$ deposited $\$F > \$G_{min} > \$G_{\emptyset}$ on **Request**, it must be the case that $\mathcal{C}_{TC}$ has $\$F - \$G_{\emptyset}$ left.

This means the total expenditure in this case will be

$$\$\mathtt{G}_{\text{req}} + \$\mathtt{G}_{\text{cncl}} + \$\mathtt{F} - (\$\mathtt{F} - \$\mathtt{G}_{\emptyset}) = \$\mathtt{G}_{\text{req}} + \$\mathtt{G}_{\text{cncl}} + \$\mathtt{G}_{\emptyset}.$$

□

# D Future Work

We plan to develop TC after its initial deployment and expect it to evolve to incorporate a number of additional features. These fall into two categories: (1) Expanding the security model to address threats outside the scope of the initial version and (2) Extending the functionality of TC. Here we briefly discuss a few of these extensions.

## D.1 Expanding TC threat model

- *Freeloading protection.* Serious concerns have arisen in the Ethereum community about "parasite contracts" that forward or resell datagrams—particularly those from fee-based data feeds [**?**]. We plan to deploy a novel mechanism in TC to address this concern. Suppose contract $\mathcal{C}_U$ involves a set of parties / users $U = \{U_i\}_{i=1}^n$. Each player $U_i$ generates an individual share $(\mathsf{sk}_i, \mathsf{pk}_i)$ of a global keypair $(\mathsf{pk}, \mathsf{sk})$, where $\mathsf{sk} = \sum_{i=1} \mathsf{sk}_i$ and $\mathsf{pk} = \prod_{i=1} \mathsf{pk}_i$, and communicates a ciphertext $E_{\mathsf{pk}_{TC}}[\mathsf{sk}_i]$ to $\mathcal{C}_{TC}$, e.g., by including it in a datagram request. Players then jointly set up under public key $\mathsf{pk}$ a wallet $\mathcal{W}^*$ for datagram transmission by $\mathcal{C}_{TC}$.

  Thanks to the homomorphic properties of ECDSA, $\mathcal{C}_{TC}$ can compute $\mathsf{sk}$ (non-interactively) and send datagrams from $\mathcal{W}^*$. But the users $U$ collaboratively *can also compute* $\mathsf{sk}$ *and send messages from* $\mathcal{W}^*$. Consequently, while each user $U_i$ can individually be assured that a datagram sent to $\mathcal{C}_U$ by $\mathcal{C}_{TC}$ from $\mathcal{W}^*$ is valid (as $P[i]$ didn't collude in its creation), other players cannot determine whether a datagram was produced by $\mathcal{C}_{TC}$ or $U$, and thus whether or not it is valid. Such a *source-equivocal datagram* renders data from parasite contracts less trustworthy and thus less attractive.

- *Traffic-analysis protection.* As noted in the body of the paper, Relay can observe the pattern of data sources accesses made by TC. By correlating with activity in $\mathcal{C}_{TC}$, an adversarial Relay can thus infer the data source targeted by private datagrams, as well as the timing—and potentially, based on traffic analysis, of the Enclave's HTTPS requests. (See, e.g., [**?**].) In the example contract FlightIns in Section **??**, this issue is partially addressed through the insertion of random delays into TC responses. We intend to develop a comprehensive approach to mitigating traffic analysis in TC. This approach will include, for the problem of traffic analysis of web scraping, the incorporation of facilities in the Enclave to make chaff or decoy data requests, i.e., false requests, to both the true data source and well as non-target data sources.

- *Revocation support.* Revocation of two forms can impact the TC service.

  First, the certificates of data sources may be revoked. To address this issue, given its ability to establish external HTTPS connections, TC could easily make use of Online Certificate Status Protocol (OCSP) certificate checking. This functionality would amount to an additional form of web scraping, and could be executed in parallel with web scraping to support datagram requests, resulting in minimal additional latency.

  Second, an SGX host could become compromised, prompting revocation of its EPID signatures by Intel. The Intel Attestation Service (IAS) will reportedly provide support for online attestation verification and thus for revocation. Conveniently, clients use the IAS when checking the attestation $\sigma_{\text{att}}$, so no modification to TC is required to support the service.

- *SLAs.* Were TC to be deployed as a fee-for-service system, requesters might wish to see Service Level Agreements (SLAs) enforced. A temporary outage on a TC host or a malicious Relay could cause a delay in datagram delivery, potentially with aftereffects in relying contracts. An SLA could be implemented as an indemnity paid to a contract if a datagram is not delivered within a specified period of time. This feature could itself be implemented, for example, as an entry point in $\mathcal{C}_{TC}$. (Naturally care would be required to ensure that $\mathcal{C}_{TC}$ holds funds sufficient for payout in the case of a general delivery failure.)

## D.2 Expanding TC functionality

- *New opcodes.* Ethereum's developers [**?**] have indicated an intention to expand the range of supported cryptographic primitives in Ethereum and stated that they are amenable to the authors' suggestion of incorporating opcodes supporting Intel's EPID in particular, which would enable efficient attestation verification within the blockchain.

- *Migration to data-source feeds.* Ultimately, we envision that data sources may wish themselves to serve as authenticated data feeds. To do so, they could simply stand up TC as a front end. As a first step along this path, however, an independent TC service might provide support for XML-labelled data from data sources, enabling more accurate and direct scraping and intentional identification of what data should be served. We

plan to build support for such explicit data labeling into TC should this approach prove attractive to data sources.

- *Generalized custom datagrams.* In our example smart contract SteamTrade, we demonstrated a custom datagram that is essentially hardwired: It employs a user's credentials to scrape her individual online account. A more general approach would be to allow contract owners to specify their own generalized functionalities–scrapers and/or confidential contract modules—as general purpose code, achieving a data-source-enriched emulation of private contracts as in Hawk [**?**], but with much lower resource requirements. Furnishing large custom datagrams on the blockchain would be prohibitively expensive, but off-chain loading of code would be quite feasible. Of course, many security and confidentiality considerations arise in a system that allows users to deploy arbitrary code, giving rise to programming language challenges that deployment of this feature in TC would need to address.

# E Implementation Pseudocode

## E.1 Application Contracts

FlightIns **blockchain contract**

**Constants**

$D_{\min}$ := minimum delay to pay out insurance
$T_{\text{flight}}$ := Town Crier flight info request type
$\$F_{\text{prem}}$ := premium to buy insurance
$\$F_{\text{pay}}$ := payout if flight is canceled or delayed
$\$F_{TC}$ := fee payed to TC for datagram delivery

**Functions**

**Init:** On recv $(\mathcal{C}_{TC})$:
Save address of $\mathcal{C}_{TC}$

**Insure:** On recv $(encFN, \$f)$ from $\mathcal{W}_U$:
Assert $\$f = \$F_{\text{prem}}$
params $:= [T_{\text{flight}}, encFN]$
callback := `this`.**Pay**
id $:= \mathcal{C}_{TC}$.**Request**(params, callback, $\$F_{TC}$)
Store $(\text{id}, \mathcal{W}_U)$

**Pay:** On recv $(\text{id}, D)$ from $\mathcal{C}_{TC}$:
Retrieve and remove stored $(\text{id}, \mathcal{W}_U)$
*// Abort if not found*
If $D \geq D_{\min}$
Send $\$F_{\text{pay}}$ to $\mathcal{W}_U$

Figure 13: The FlightIns application contract

CashSettledPut **blockchain contract**

**Constants**

$T_{\text{stock}}$ := Town Crier stock info request type
$\$F_{TC}$ := fee payed to TC for datagram delivery

**Functions**

**Init:** On recv $(\mathcal{C}_{TC}, \text{ticker}, P_S, P_U, M, \text{expr}, \$f)$ from $\mathcal{W}_{\text{issuer}}$
Assert $\$f = (P_S - P_U) \cdot M + \$F_{TC}$
Save all inputs and $\mathcal{W}_{\text{issuer}}$ to storage.

**Buy:** On recv $(X, \$f)$ from $\mathcal{W}_U$:
Assert isRecovered not set
and timestamp $<$ expr
and $\mathcal{W}_{\text{buyer}}$ not set
and $X \leq M$
and $\$f = (X \cdot P_U)$
Set $\mathcal{W}_{\text{buyer}} = \mathcal{W}_U$
Save $X$ to storage
Send $(P_S - P_U)(M - X)$ to $\mathcal{W}_{\text{issuer}}$
*// Hold* $P_S \cdot X + \$F_{TC}$

**Put:** On recv () from $\mathcal{W}_{\text{buyer}}$:
and timestamp $<$ expr
and isPut not set
Set isPut
params $:= [T_{\text{stock}}, \text{ticker}]$
callback := `this`.**Settle**
$\mathcal{C}_{TC}$.**Request**(params, callback, $\$F_{TC}$)

**Settle:** On recv $(\text{id}, P)$ from $\mathcal{C}_{TC}$:
If $P \geq P_S$
Send $P_S \cdot X$ to $\mathcal{W}_{\text{issuer}}$
Return
Send $(P_S - P)X$ to $\mathcal{W}_{\text{buyer}}$
Send all money in contract to $\mathcal{W}_{\text{issuer}}$
Send $P \cdot X$ to $\mathcal{W}_{\text{issuer}}$

**Recover:** On recv () from $\mathcal{W}_{\text{issuer}}$:
and isPut not set
and isRecovered not set
and ($\mathcal{W}_{\text{buyer}}$ not set
or timestamp $\geq$ expr)
Set isRecovered
Send all money in contract to $\mathcal{W}_{\text{issuer}}$

Figure 14: The CashSettledPut application contract