

# CS 453/698: Software and Systems Security

## **Module: Non-technical aspects of Security**

Lecture: Ethics, Legal Issues, Regulation, Compliance

Adam Caulfield

*University of Waterloo*

Spring 2025

# Reminders & Recap

## Reminders:

- [A4 is released](#)
  - Due tomorrow!

## Recap – last time we covered:

Side-channel attacks

# Outline

## **Ethics**

- Intro to ethics and ethical
- Computer security Trolley Problems

## **Legal Issues and Regulation**

- What are legal problems that relate to security/privacy?
- What are regulations related to security/privacy issues?

## **Compliance:**

- How to demonstrate adherence with regulation?
- A few example methods
- Ongoing industry efforts

# Ethics

## What are ethics?

***Definition:*** *moral principles that govern a person's behavior or the conducting of an activity;*

***The analysis, evaluation, and promotion of a correct conduct according to a particular standard***

The “particular standard” → ethical framework

# Ethical Dilemmas

**Debate ethical frameworks through proposing dilemmas.**

Dilemmas:

- Proposes two options
- Both contain some undesired outcome
- Used to show that moral intuitions will most likely diverge
- Different ethical frameworks present different answers

# Ethical Dilemmas

Example: the classic Trolley Problem ([plus even more](#))

## A Classic Dilemma: The Trolley Problem

### Context:

- A runaway trolley with no brakes is heading straight along a set of tracks.
- Five people are tied to those tracks.
- One person is tied to an alternate set of tracks.
- A trolley operator has the ability to change the trolley's path and make it head down the alternate set of tracks.

### The choice for the trolley operator:

- *Do nothing:* Five people die.
- *Make the trolley take the alternate set of tracks:* One person dies.

# Ethical Frameworks

## **What are ethical frameworks?**

Define different approaches for reasoning about the morality of a particular action.

Main ethical frameworks:

- Consequentialist ethics
- Deontological ethics
- Virtue ethics
- Discourse ethics

# Ethical Frameworks

## Consequentialist Ethics

**Definition:** An ethical framework that evaluates actions based on their ***outcomes or consequences***

Key principle: “the ends justify the means”

If a certain action prevents greater harm in the future, despite causing minor harm in the present, it may be justified under consequentialism



# Ethical Frameworks

## Deontological Ethics

**Definition:** An ethical framework that evaluates actions based on whether they are ***aligned with a set of rules or duties***, regardless of outcomes

Key principle: “do the right thing no matter the consequence”

Under deontological ethics, actions that are deemed wrong should never be taken (e.g., violating autonomy, destroying property) even if it leads to positive consequences

# Ethical Frameworks

## Virtue Ethics

**Definition:** An ethical framework that focuses on the moral character of the agent rather than on particular rules or consequences

Key principle: “act as a virtuous individual would act”

Under virtue ethics, the rules or outcomes are not considered, but what decision would allow an actor/agent to maintain their virtue (e.g., honesty, responsibility, etc.)

# Ethical Frameworks

## Discourse Ethics

**Definition:** An ethical framework believes ethical truths emerge through rational and inclusive dialogue among affected parties, rather than a single factor.

Key principle: “morality is determined by what can be agreed upon”

Under discourse ethics, morality of a decision should only be concluded after debate among all stakeholders (e.g., users, policymakers, developers).

# Real world examples

## Edward Snowden NSA whistleblower (2013)

- Snowden leaked classified NSA documents revealing mass surveillance
- NSA → operating deontological justification
  - Surveillance without disclosure was due to patriotic duty for national security
- Snowden → operating on consequential & virtue ethics
  - Believed harm of secrecy outweighed harm of disclosure
  - People have the right to know if their rights are being violated

## Facebook whistleblower (2021)

- Frances Haugen leaked Facebook internal research
- Revealed Facebook wasn't taking hateful/divisive content seriously
- Haugen → operating on virtue & discourse ethics
  - Informing the public was the right thing to do
- Facebook → operating on consequential ethics
  - Prioritizing the bottom line

# Computer Security Trolley Problems

## Three hypothetical computer security ethical dilemmas

- Medical device vulnerability
- Study of stolen data
- Inadvertent disclosure
- [From this work](#)

## Let's propose an approach from the framework of:

- Consequentialist ethics
- Deontological ethics
- Virtue ethics
- Discourse ethics

# Scenario 1: Medical Device Vulnerability

## Context:

- Researchers discover a serious vulnerability in wireless implantable medical devices
- The manufacturer no longer exists, and the device cannot be patched/updated
- The device is still used by existing patients and being implanted in new ones
- There is no realistic chance of exploiting the vulnerability, and disclosure offers no technical or field-wide benefits
- Public disclosure could cause harm (e.g., patient fear, risky device removals)
- Lack of disclosure denies patients informed consent

## Choice:

- **Option A:** disclose the vulnerability, respecting patient autonomy
- **Option B:** do not disclose, prioritizing patient safety but withholding knowledge

# Scenario 2: Study of stolen data

## Context:

- Company B's AI hiring system is suspected of racial and gender bias and potentially has vulnerabilities
- Hackers steal all internal data and models that are in use, and post them online
- Researchers download them before the public copies are removed
- This data includes sensitive applicant information, data about the models in use, job matching outputs
- Applicants publicly request the data to be deleted
- The researchers want to study the data to evaluate bias, security issues, and to propose improvements that prevent future harm
- Studying the data requires keeping a copy long-term

## Choice:

- **Option A:** study the data, advancing public interest in fairness of AI hiring
- **Option B:** do not study the data, respecting the rights and wishes of applicants

# Scenario 3: Inadvertent disclosure

## Context:

- Company C's employee is serving on a program committee to conduct peer review of current computer security research
- The employee reads a submission that discloses a serious, unpublished vulnerability in their company's product
- The program committee requires strict confidentiality, which the employee agreed to with their company's knowledge and approval
- The vulnerability is severe and will take time to patch, posing a potential risk to users if left unaddressed

## Choice:

- **Option A:** break confidentiality agreement to disclose the vulnerability
- **Option B:** respect the confidentiality agreement and wait to disclose



# Computer Security Ethical problems

**Many ethical problems can arise in the world of computer security**

Vulnerability disclosure

- When to disclose? Publicize exploits?

Privacy vs. security

- To what extent should user privacy be violated if it leads to more security?

Privacy vs. transparency

- How to provide transparency, if it might risk privacy?

Allocation of security resources

- Which components of the system should be protected?

# Legal Issues and Regulation

## **Common Cybersecurity-related crimes**

### Unauthorized access (aka hacking)

- Accessing systems without permissions

### Data theft and breaches

- Inducing or enabling stealing or leaking of sensitive information

### Ransomware

- Encrypting victims data and demanding payment

### Denial of service attacks

- Mirai botnet

### Unauthorized usage

- Intellectual property
- Copyrighted content

# Legal Issues and Regulation

## **Prosecution is difficult**

### Attribution is difficult

- Attackers can hide their identity
- Operate through global botnets

### Cross-border jurisdiction issues

- Locations of attackers & victims may vary
- Infrastructure may span multiple countries

### Dual-use technologies

- Tools that can be used for testing (e.g., nmap, metasploit, etc) can also be used maliciously

# Legal Issues and Regulation

## **Ethical hacking**

Probing systems for vulnerabilities to improve security, done in good faith and often with permission

### Types:

- White box vs. black box

### Legality:

- When given explicit authorization
- E.g., penetration testing contracts, bug bounty programs, etc.

### Gray areas:

- Without permission can violate laws
- Responsible disclosure (race condition)
- Companies can still sue

# Legal Issues and Regulation

## **EU General Data Protection Regulation (GDPR)**

- Applies to any organization processing personal data
- Requires lawful basis for processing data
  - Consent, contract, legal obligations, etc.
- Enforces data subject rights
  - Access, rectification, erasure (right to be forgotten), and more
- Mandates assessments for high-risk processing
  - E.g., profiling or surveillance
  - Data Protection Impact Assessments
- “Privacy by design and by default”
- Requires notification of data breaches within 72 hours to regulators
- [Read more](#)

# Legal Issues and Regulation

## **United States - California Consumer Privacy Act (CCPA) / California Privacy Rights Act (CPRA)**

- Applies to businesses that meet certain thresholds
  - \$25M or more in revenue
  - Having data on 100K or more individuals
- Grants consumers rights to access, delete, and correct their information
- Provides right to opt out of sale/sharing of personal data
  - Opt In for children under 16
- Requires “notice at collection”
- Enables civil penalties for violations, including \$7.5K per intentional violation
- [Read more](#)

# Legal Issues and Regulation

## **Canada Personal Information Protection and Electronic Documents Act (PIPEDA)**

- Federal policy: applies to private-sector organizations engaged in commercial activity across provinces
- Requires meaningful consent for collection, use, disclosure of personal information
  - PI must be collected for explicit and reasonable purposes
- Grants individuals rights to access and correct their personal information
- Organizations must adopt safeguards appropriate to the sensitive data
- Must notify affected individuals of breaches
- [Read more](#)

# Legal Issues and Regulation

## Ontario's Bill 194:

- Strengthening Cyber Security and Building Trust in the Public Sector Act, 2024
- Empowers government to mandate incident reporting, standards, oversight
- Regulates public sector AI use
  - Risk assessments, public transparency, human oversight
- Digital technology protections for minors
- Requires breach notifications
- Establishes whistleblower protections
- And more



# Legal Issues and Regulation

## Emerging regulations → AI regulations

### EU AI act (2024)

- First comprehensive AI regulation passed globally
- Requires transparency & human oversight for high risk cases
- Bans certain “unacceptable risk”
  - Social scoring AI, biometric identification/classification, facial recognition in public
- Transparency
  - Disclosing that content was generated by AI
  - Designing model to prevent generating illegal content
  - Publishing summaries of copyrighted data
- [Read more](#)

### Others

- [Canada Artificial Intelligence and Data Act](#)
- Global governance trends → [United Nations](#), [BRICS+ countries](#)

# Compliance

## How to comply with regulations?

Well a few steps are required:

- Implement a system that is indeed complying with the regulation
- Then add mechanisms to either
  - Enforce a certain policy
  - Prove that your system is indeed complying
- Requires some cryptography or security mechanisms from this class!

# Compliance

## **Digital Rights Management (DRM)**

- A form of digital copyright content protection
- Involves enforcement of security policy related to digital rights of a content in a particular software product
- Specified by the owner

## Key Elements of DRM

- Content owner + customer (with DRM client)
- Content is encrypted and can only be decrypted based on
  - Device-bound key
  - Usage check with license
- Difficult to obtain: a lot of security requirements

# Compliance

## **Proof of data erasure:**

How to prove certain data from erased from memory?

- Need to report on some measurement memory
- Data (D) erased → memory has some configuration that is not D

## Proof of data erasure:

How to prove certain data from erased from memory?

- Need to report on some measurement memory
- Data (D) erased  $\rightarrow$  memory has some configuration that is not D

### Secure Code Update for Embedded Devices via Proofs of Secure Erasure\*

Daniele Perito<sup>1</sup> and Gene Tsudik<sup>2</sup>

<sup>1</sup> INRIA Rhône-Alpes, France

<sup>2</sup> University of California, Irvine, USA

# Compliance

## Proofs of Secure Erasure:

### Model:

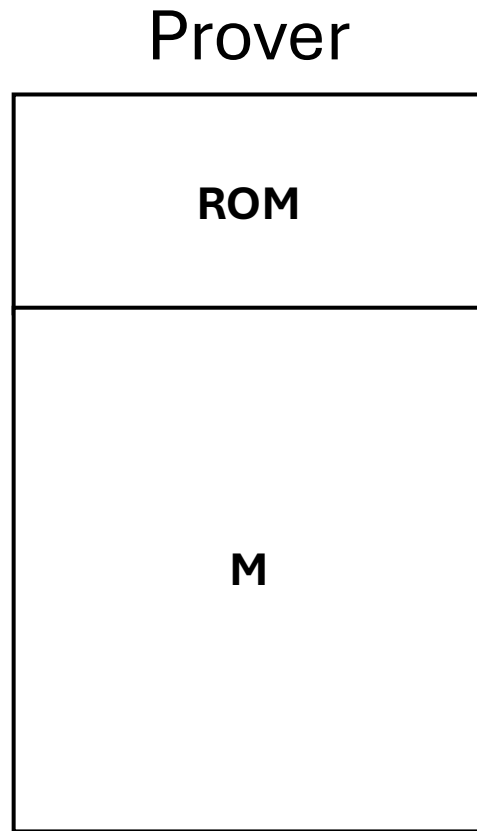
- Verifier (V)  $\rightarrow$  request P to prove it erased data
- Prover (P)  $\rightarrow$  simple embedded device
- Prover has writeable memory (M) of size  $n$

### Assumptions:

- Software adversary
- Prover has small amount of ROM

# Compliance

## Proofs of Secure Erasure:

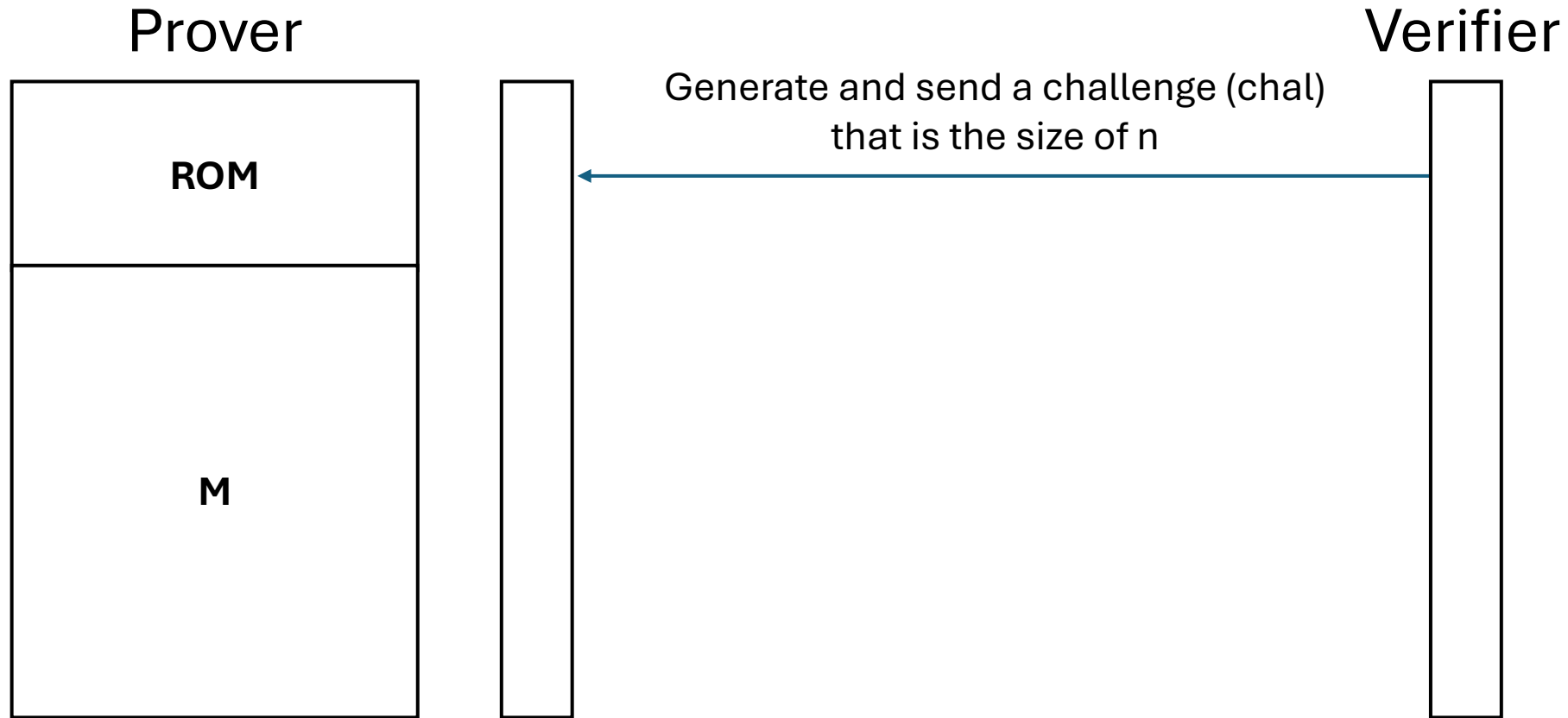


Verifier



# Compliance

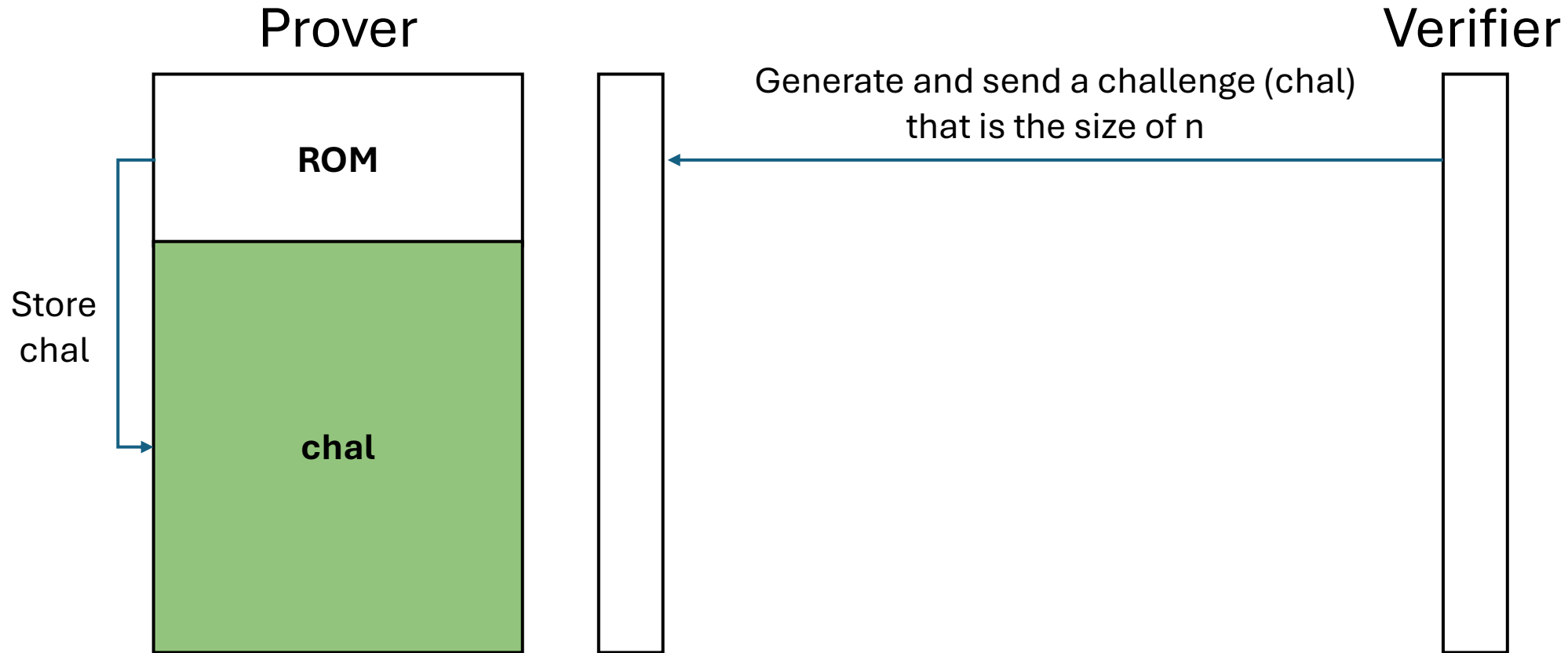
## Proofs of Secure Erasure:





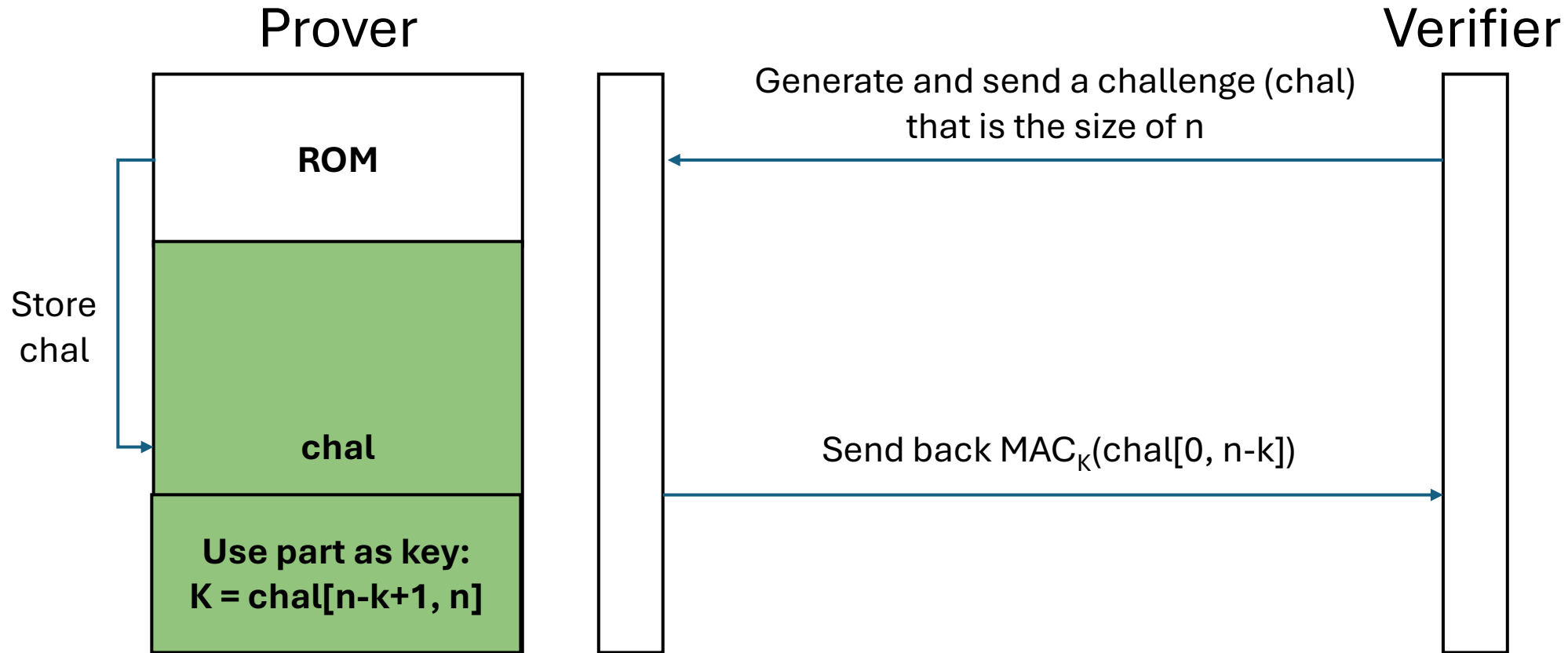
# Compliance

## Proofs of Secure Erasure:



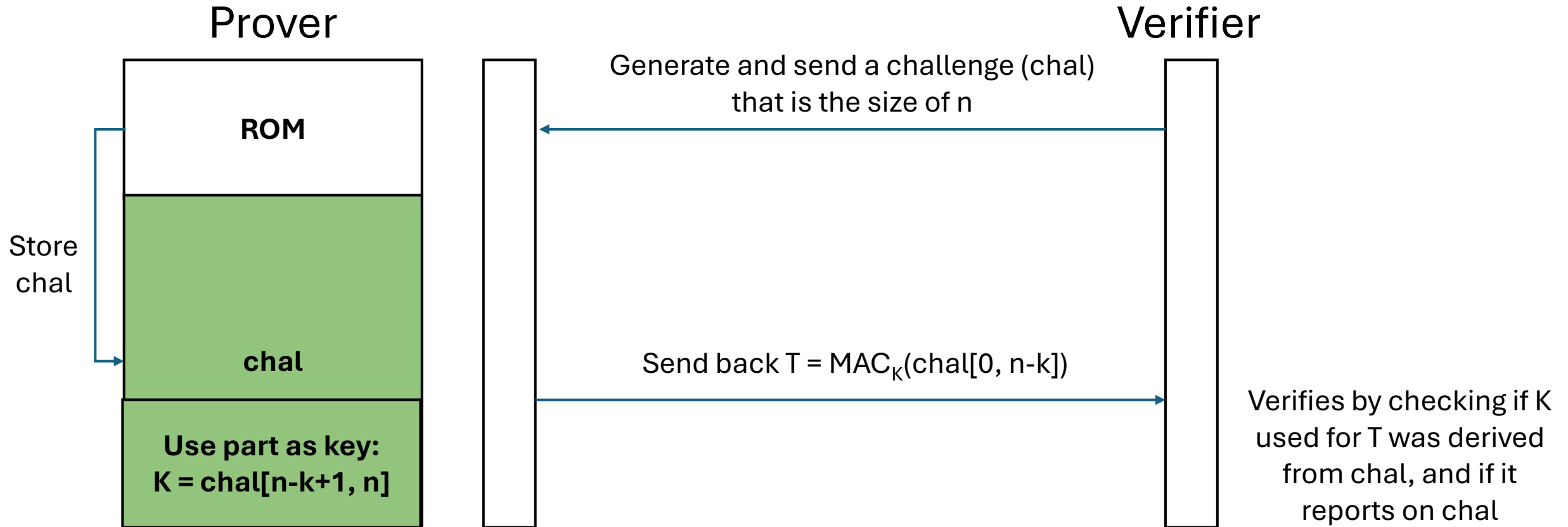
# Compliance

## Proofs of Secure Erasure:



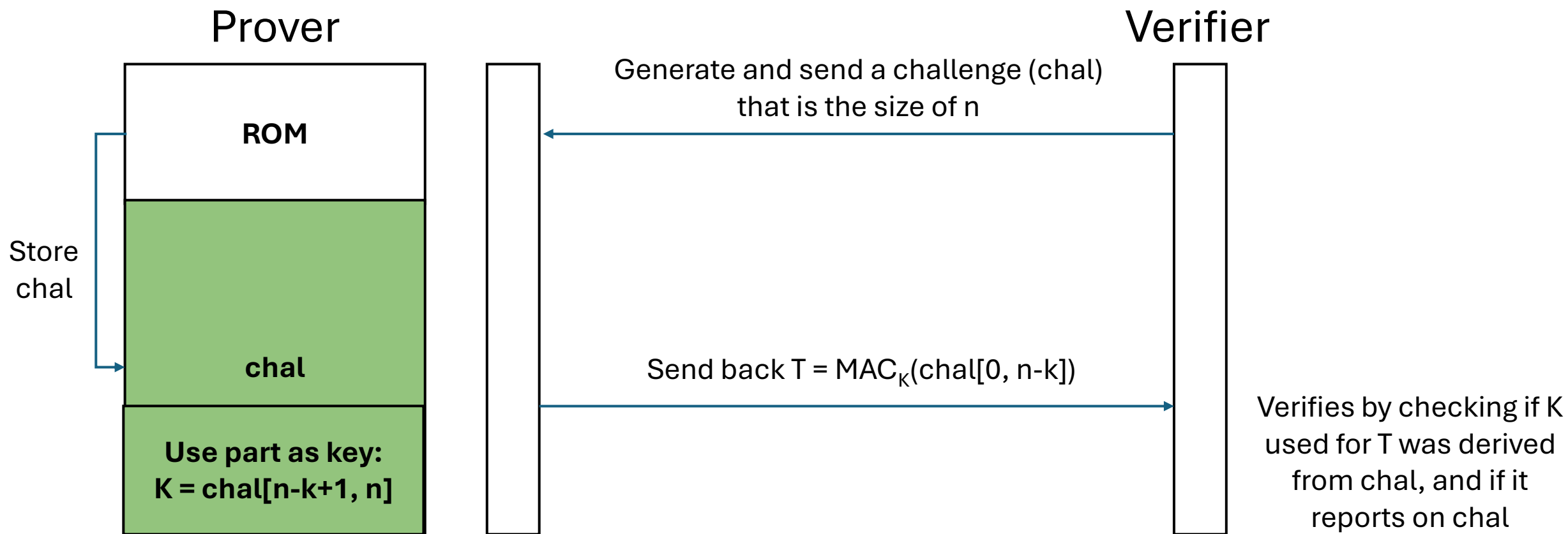
# Compliance

## Proofs of Secure Erasure:



# Compliance

## Proofs of Secure Erasure:



What other assumptions are in place?

# Compliance

## Proofs of Secure Erasure:

### Model:

- Verifier (V)  $\rightarrow$  request P to prove it erased data
- Prover (P)  $\rightarrow$  simple embedded device
- Prover has writeable memory (M) of size  $n$

### Assumptions:

- Software adversary
- Prover has small amount of ROM
- **No DMA, peripherals that interfere with computations**
- **No prior knowledge of challenge**
- **Adv. Prover cannot consult external devices**
- **Must completely store before measuring**

**Proofs of Unlearning:** A little more comple, relies on TEEs (Intel SGX)

Applied to ML:

- Given that model (M) was trained on some data D
- Generate proof that D was “unlearned” from the model (M)

IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 19, 2024

3309

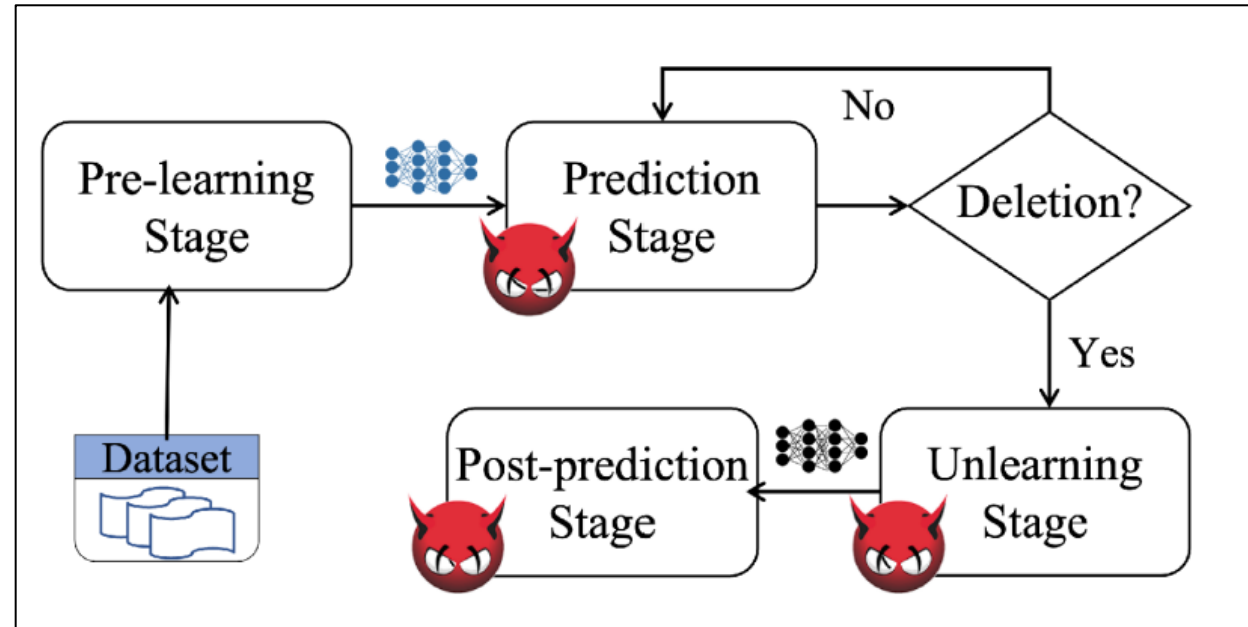
## Proof of Unlearning: Definitions and Instantiation

Jiasi Weng<sup>id</sup>, *Member, IEEE*, Shenglong Yao, Yuefeng Du<sup>id</sup>, *Member, IEEE*, Junjie Huang,  
Jian Weng<sup>id</sup>, *Senior Member, IEEE*, and Cong Wang<sup>id</sup>, *Fellow, IEEE*

# Compliance

## Proofs of Unlearning:

Threat and System model:



Requires ability to:

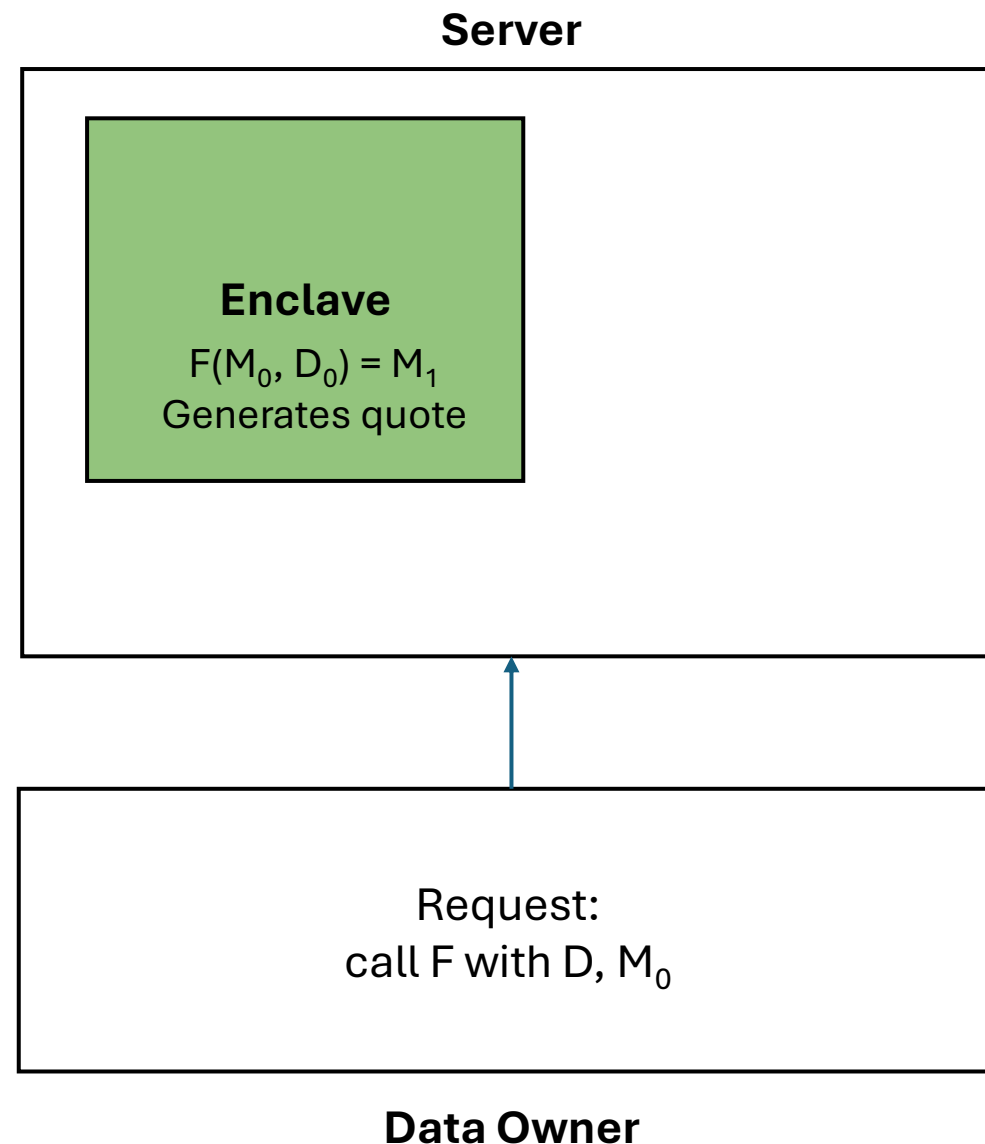
- Measure inputs and outputs of each stage
- Securely maintain them until needed

# Compliance

## Proofs of Unlearning:

Setup phase:

- Data owner uploads authenticated data ( $D_0$ )
- Data owner specifies a learning algorithm ( $F$ ) to use and initial model ( $M_0$ )
- Server “learns” a new model  $M_1$  from  $F(D_0, M_0)$
- Generates SGX quote to capture process



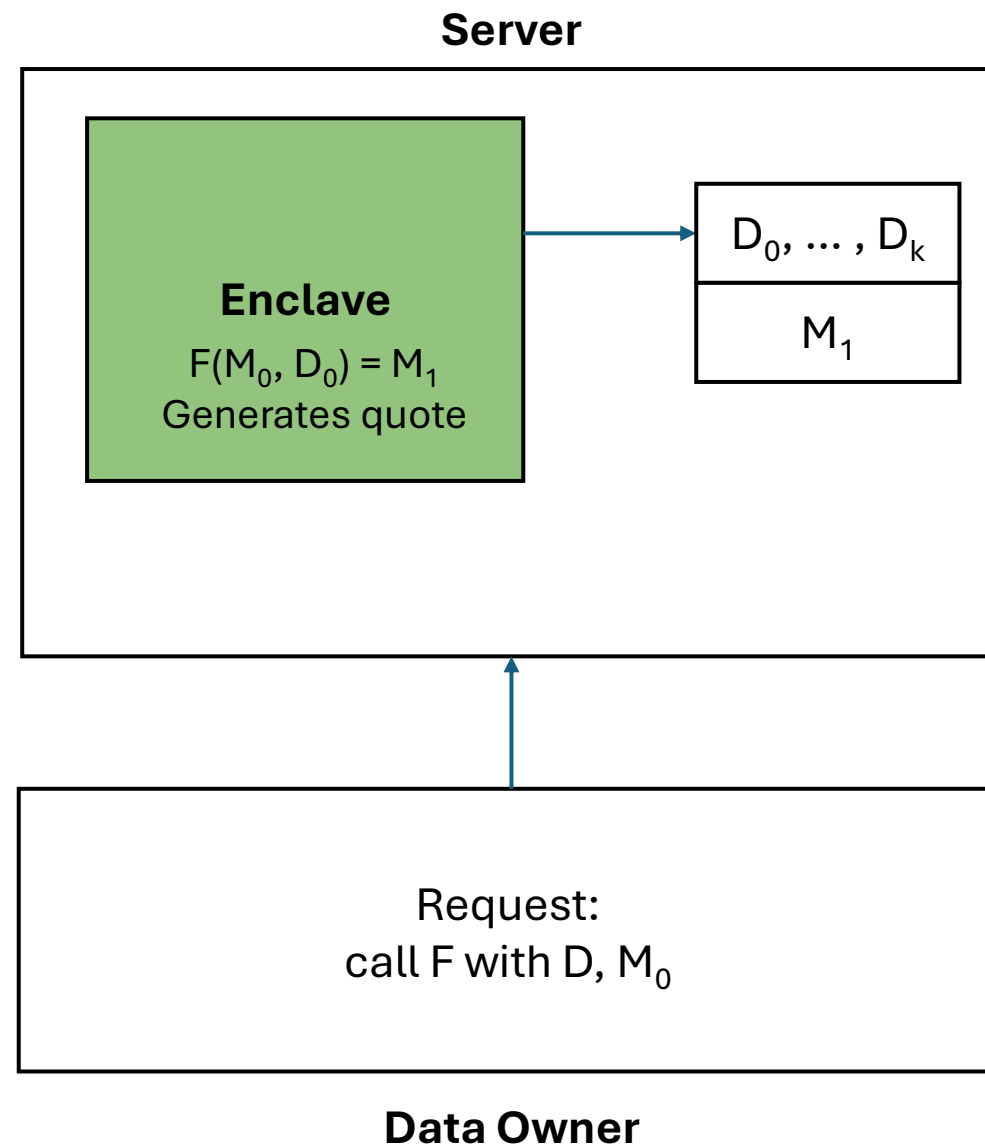


# Compliance

## Proofs of Unlearning:

Setup phase:

- Data owner uploads authenticated data ( $D_0$ )
- Data owner specifies a learning algorithm ( $F$ ) to use and initial model ( $M_0$ )
- Server “learns” a new model  $M_1$  from  $F(D_0, M_0)$
- Generates SGX quote to capture process
- Signs and stores externally

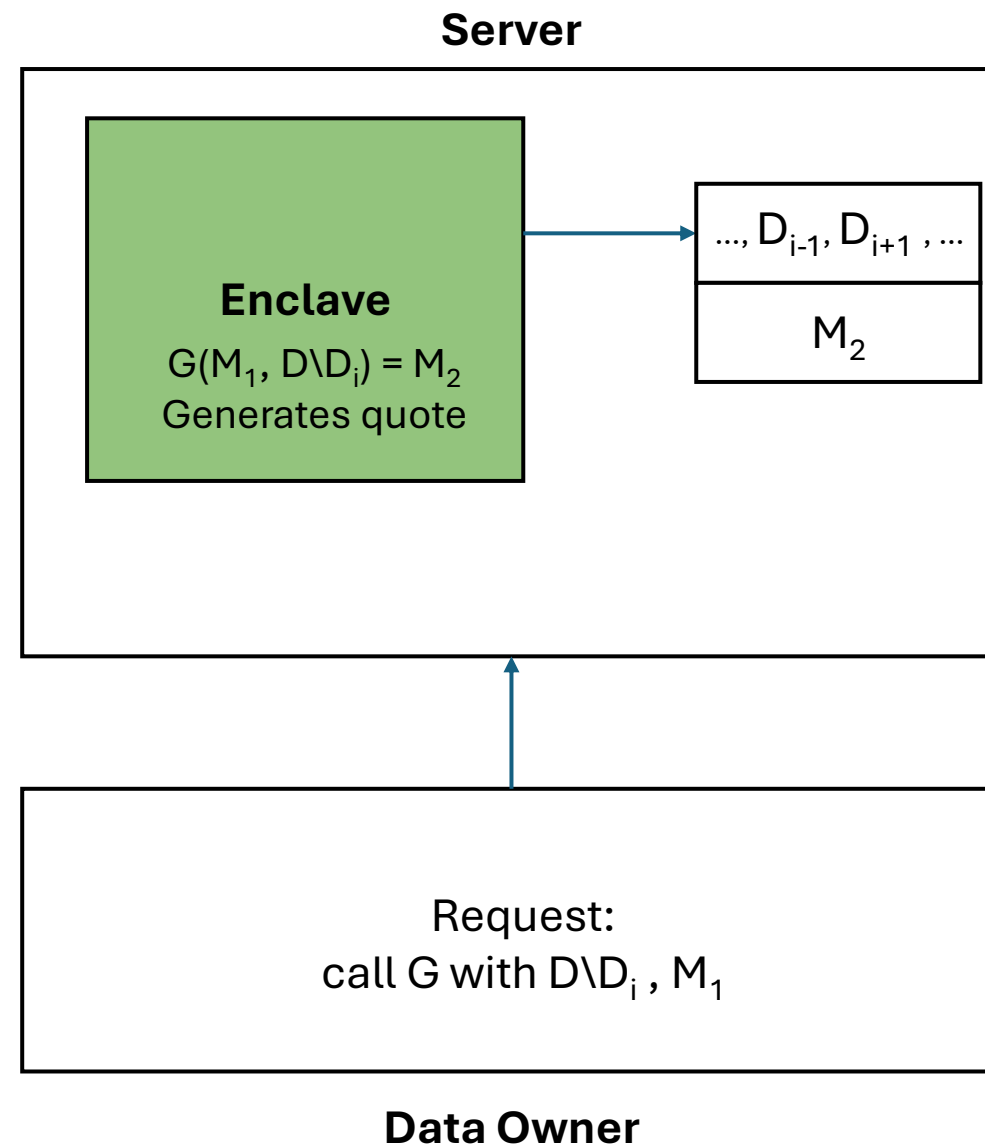


# Compliance

## Proofs of Unlearning:

Deletion phase:

- Delete  $D_i$  from storage and unlearn from  $M_1$
- Execute unlearning process  $G$  with  $D \setminus D_i$  to produce  $M_2$
- Generate quote capturing the process
- Securely store components
- Signs and stores externally



# Compliance

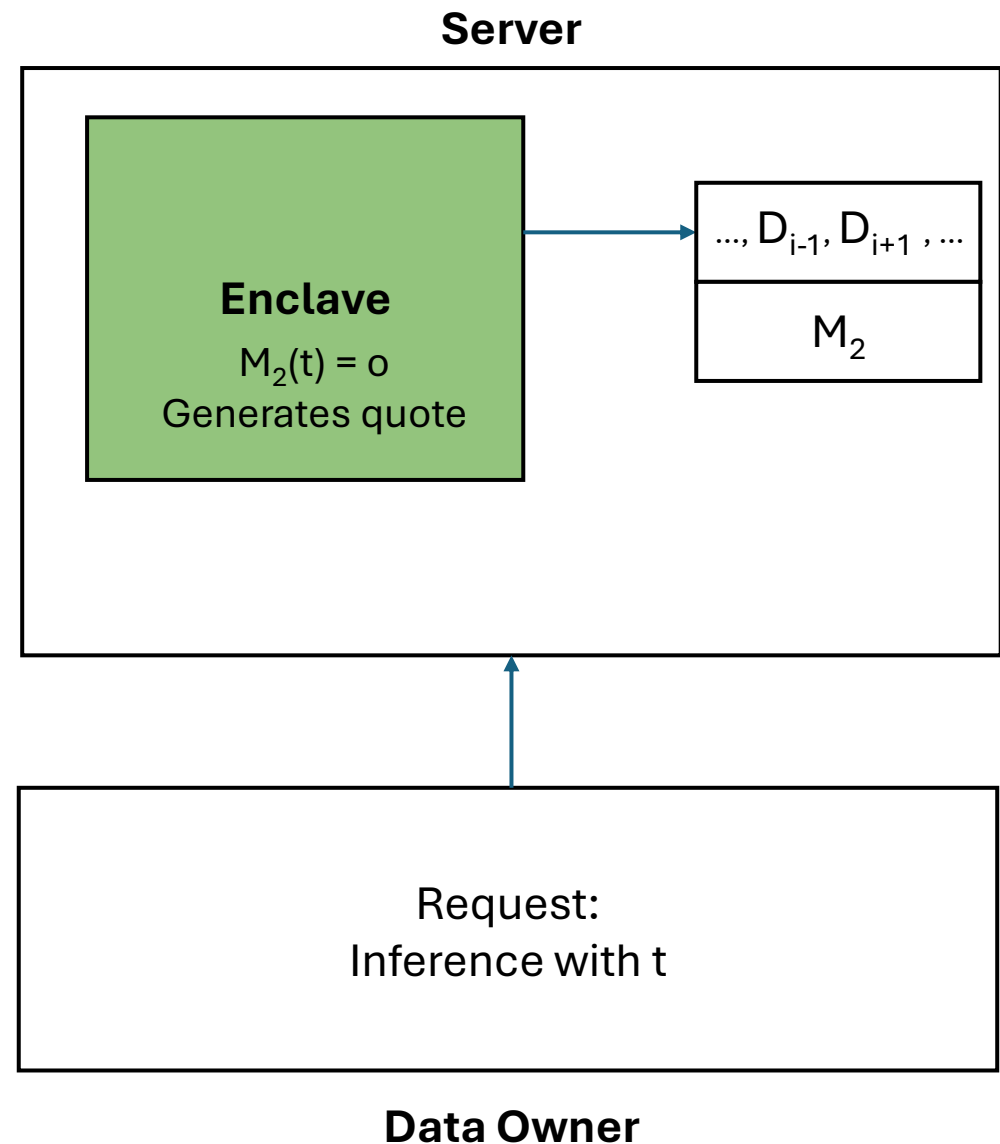
## Proofs of Unlearning:

Deletion phase:

- Delete  $D_i$  from storage and unlearn from  $M_1$
- Execute unlearning process  $G$  with  $D \setminus D_i$  to produce  $M_2$
- Generate quote capturing the process
- Securely store components
- Signs and stores externally

Inference time proof:

- Data Owner sends test data  $t$
- Server should send back  $M_2(t) = o$
- Generate quote to prove  $M_2$  was used



# Compliance

## Proofs of Unlearning:

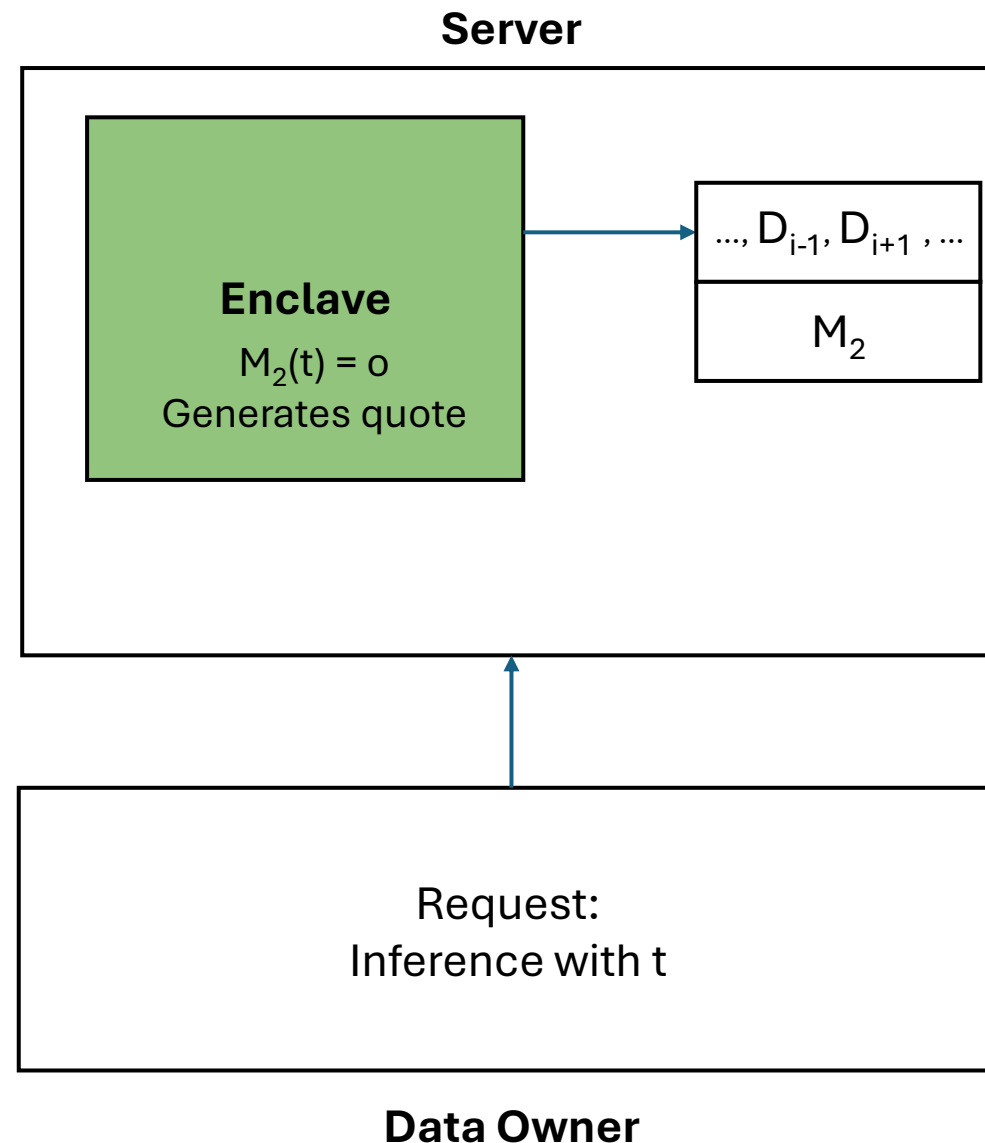
Deletion phase:

- Delete  $D_i$  from storage and unlearn from  $M_1$
- Execute unlearning process  $G$  with  $D \setminus D_i$  to produce  $M_2$
- Generate quote capturing the process
- Securely store components
- Signs and stores externally

Inference time proof:

- Data Owner sends test data  $t$
- Server should send back  $M_2(t) = o$
- Generate quote to prove  $M_2$  was used

**Assumptions?**



# Compliance

## Proofs of representative training data in ML:

- Some techniques prepare for regulation particularly related to dataset characteristics
- Some examples:
  - Doesn't include personal information
  - It has an acceptable distribution of sensitive attributes
  - E.g., fair w.r.t. gender, race, other classifications
- How to prove within a broader ML system?
  - Prove that dataset has desirable distribution
  - Prove that trained models used that dataset

## Proofs of representative training data in ML:

Laminator: using Intel SGX for Verifiable ML Property Cards

### **LAMINATOR: Verifiable ML Property Cards using Hardware-assisted Attestations**

Vasisht Duddu  
University of Waterloo  
Canada  
vasisht.duddu@uwaterloo.ca

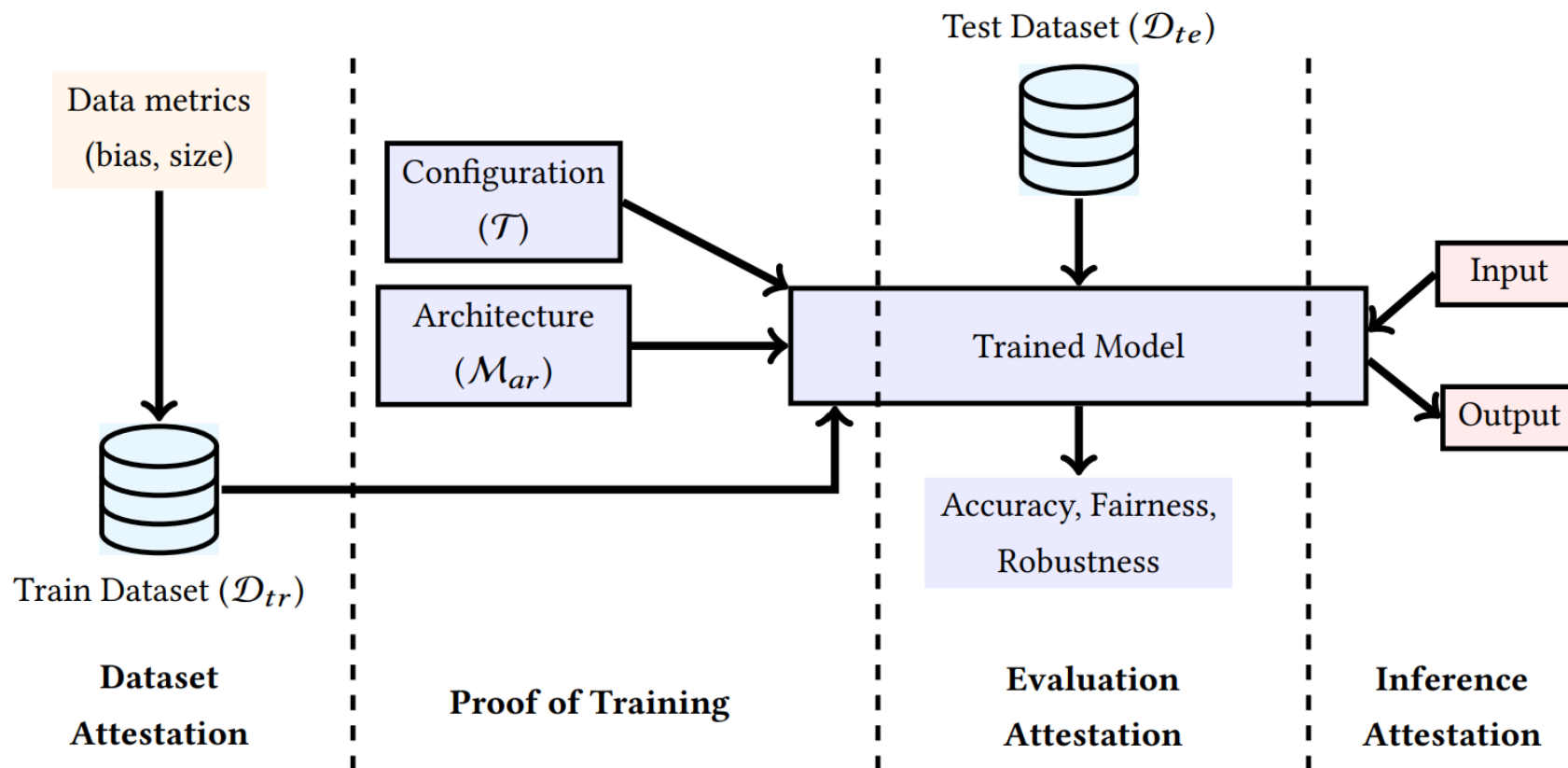
Oskari Järvinen  
Aalto University  
Finland  
oskari.jarvinen@aalto.fi

Lachlan J. Gunn  
Aalto University  
Finland  
lachlan@gunn.ee

N. Asokan  
University of Waterloo  
Canada  
asokan@acm.org

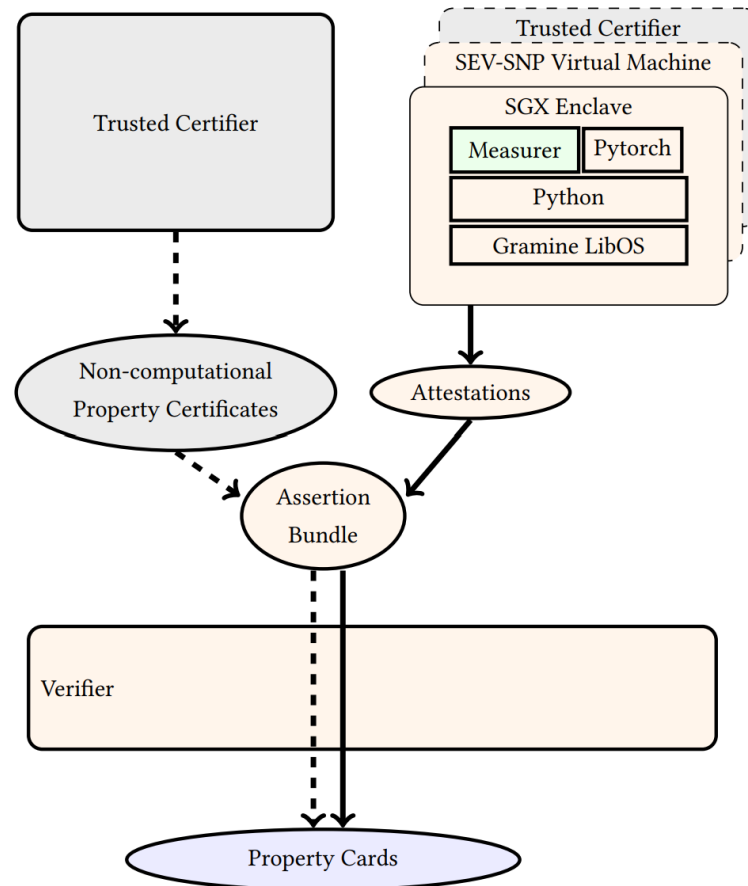
# Compliance

## Proofs of representative training data in ML:



# Compliance

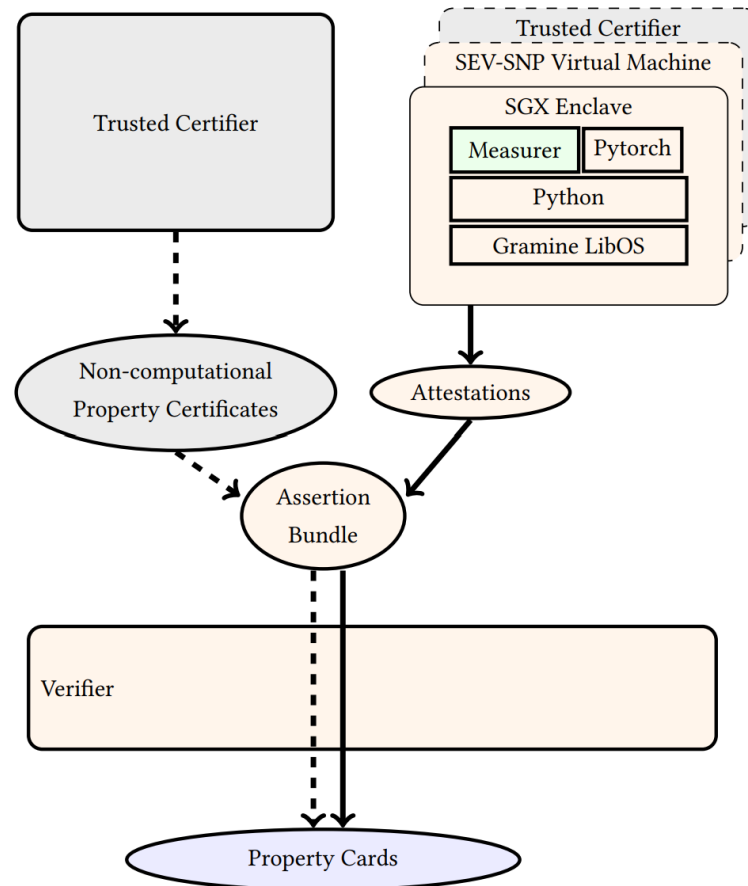
## Proofs of representative training data in ML:





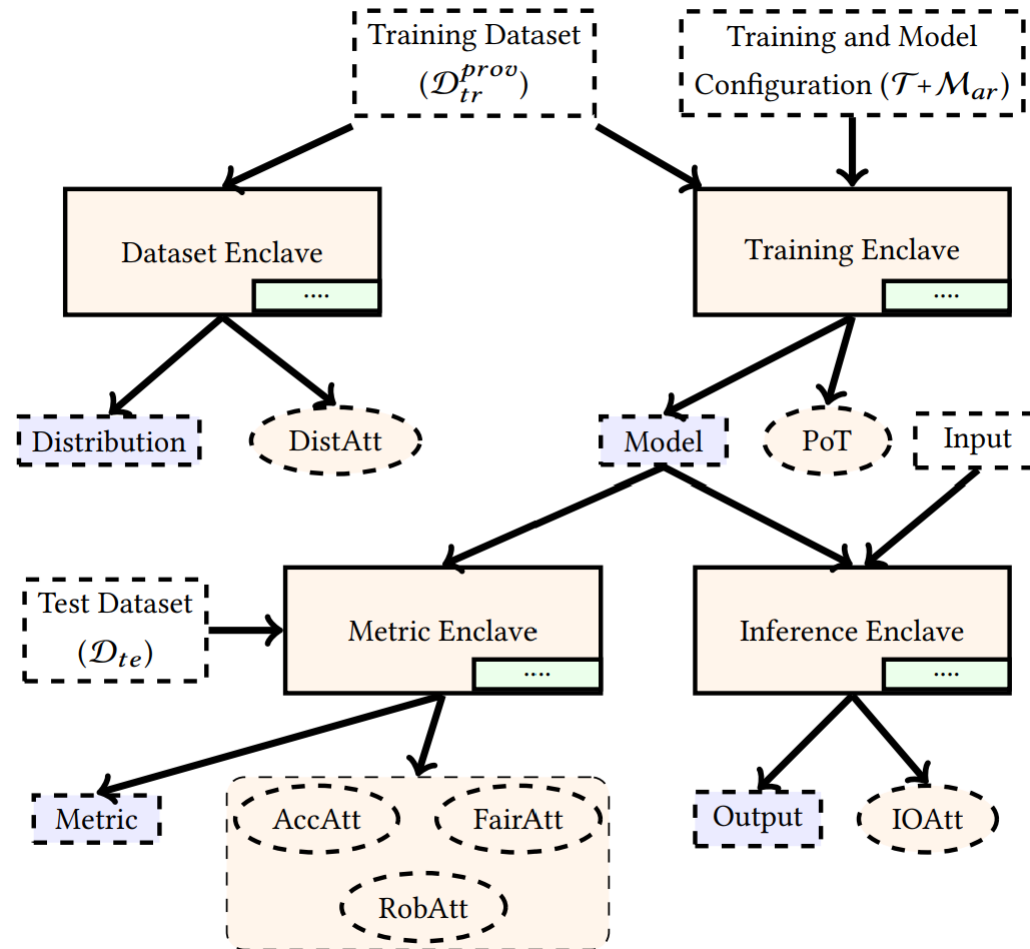
# Compliance

## Proofs of representative training data in ML:




# Compliance

## Proofs of representative training data in ML:




## Proofs of content generation & transformation

### C2PA: Coalition for Content Provenance and Authenticity


**C2PA**

Coalition for  
Content Provenance  
and Authenticity

[About](#) [FAQ](#) [Specification](#) [Content Credentials](#) [Conformance](#) [Membership](#) [Contact](#) [in](#) 

## Advancing digital content transparency and authenticity

The Coalition for Content Provenance and Authenticity, or C2PA, provides an open technical standard for publishers, creators and consumers to establish the origin and edits of digital content. It's called Content Credentials, and it ensures content complies with standards as the digital ecosystem evolves.



# Compliance

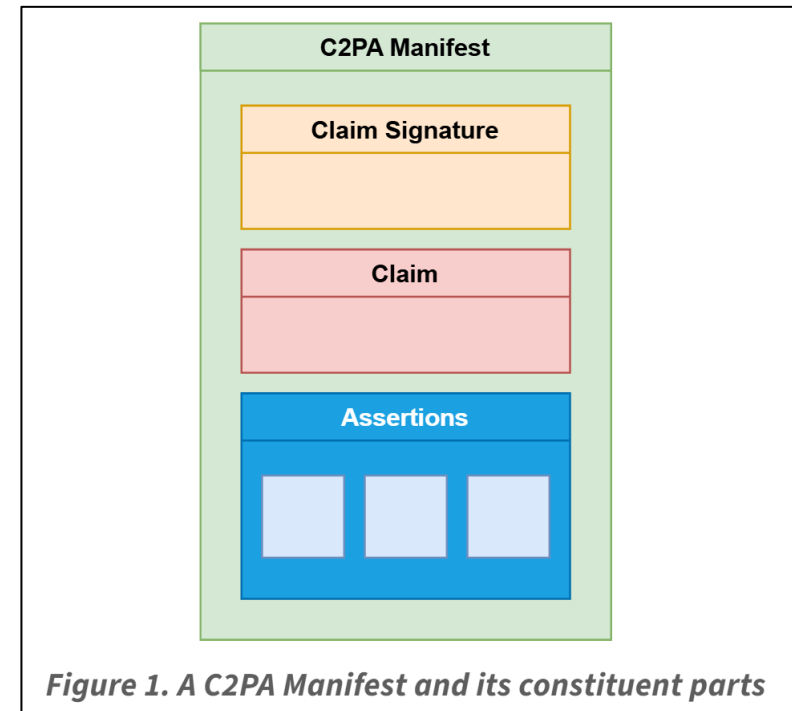
## Proofs of content generation & transformation

### C2PA: Coalition for Content Provenance and Authenticity

- Enable development of applications that prioritize proving provenance

#### Propose a specification:

- C2PA Manifest
- Signature (using device bound key)
- Assertion
  - Statement asserted by an actor
- Claim
  - Evidence of the assertions



# Compliance

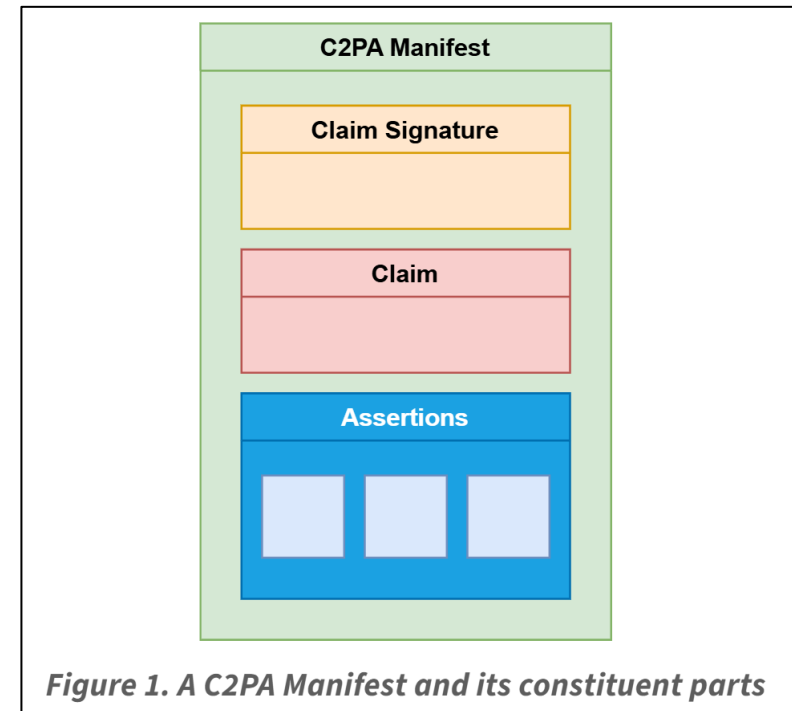
## Proofs of content generation & transformation

### C2PA: Coalition for Content Provenance and Authenticity

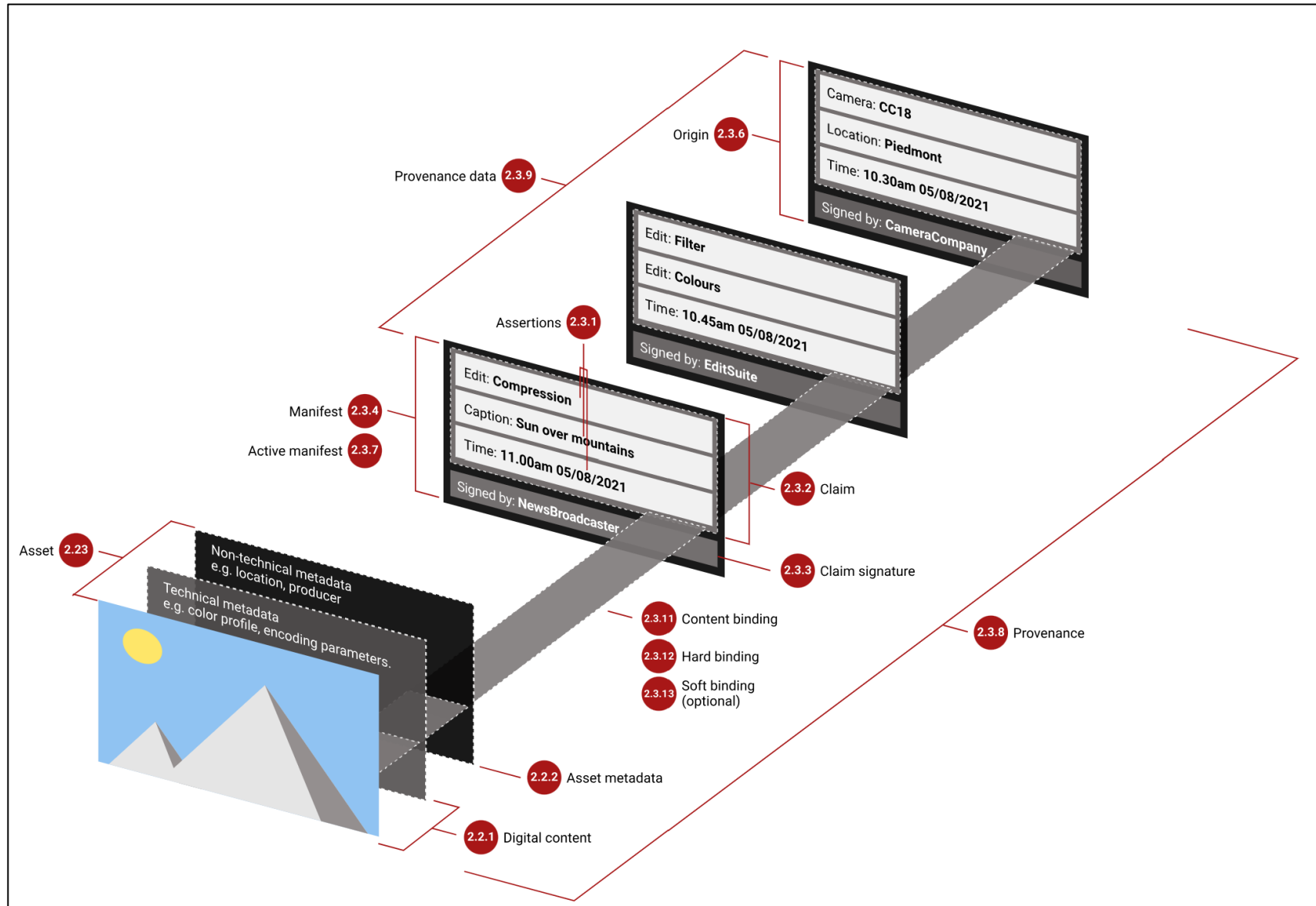
- Enable development of applications that prioritize proving provenance

#### Propose a specification:

- C2PA Manifest
- Signature (using device bound key)
- Assertion
  - Statement asserted by an actor
- Claim
  - Evidence of the assertions



# Compliance



# Compliance

## Proofs of content generation & transformation

Methods for compliance are an active area of research

- Academia
- Industry

C2PA:

Meet the Steering Committee members



# That's all for today!

## **Coming up....**

- Research related to the topics of the course
- Conclude the course

## **Reminders:**

- [A4 is due tomorrow!](#)



# That's all for today!

## Resources:

- [Computer Security Trolley Problems](#)
- Regulation
  - [GPPR](#), [CCPA](#), [PIPEDA](#), [Ontario Bill 194](#), [EU AI Act](#), [Canada AI & Data Act](#), [United Nations AI Advisory Board](#), [BRICS](#)
- Ethical dilemmas related to computer and/or security
  - [Snowden + NSA](#)
  - [Haugen + Facebook](#)
- Compliance methods
  - [C2PA](#)
  - [Laminator](#)
  - [Proofs of Unlearning](#)
  - [Proof of Update for Embedded Devices via Proofs of Secure Erasure](#)

