MIT 6.875/18.425

Foundations of Cryptography
Lecture 5

Course website: ***https://mit6875.github.io/***

TODAY

Applications of Pseudo-Random Functions (PRF):

a. Identification Protocols

b. Authentication

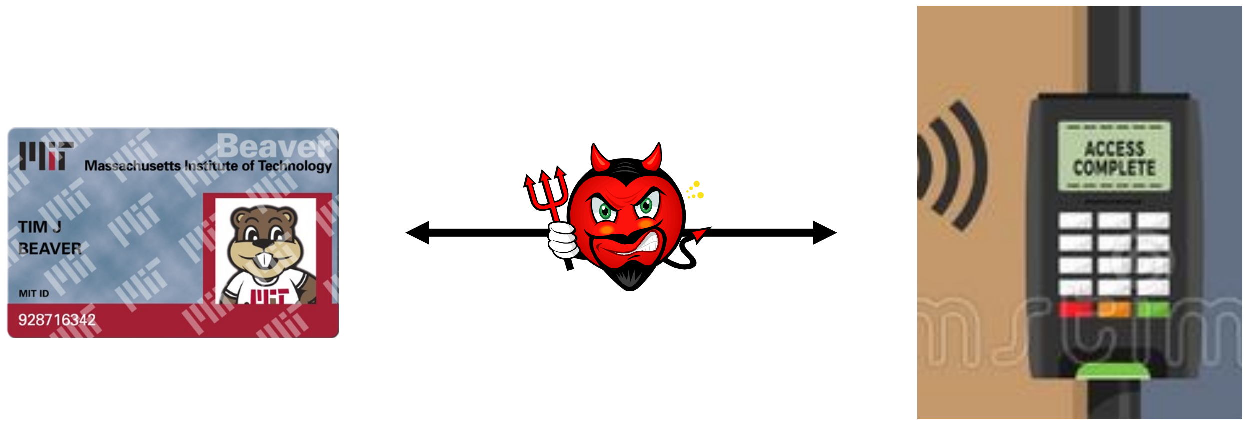
c. Encryption secure against Active Attacks

d. Applications to Learning Theory

Logistics:

- Problem Set 1 is due today at 11:59:59pm.
- Remember that you have 10 late days for this class, and you may use up to 5 for any one problem set.

Friend-or-Foe Identification



- **Adversary**: person-in-the-middle.
- Can listen to / modify the communications. Wants to impersonate Tim.

A Simple Lemma about Unpredictability

Let $f_s: \{0,1\}^\ell \to \{0,1\}^m$ be a pseudorandom function.

- Consider an adversary who requests and obtains $f_s(x_1), \ldots, f_s(x_q)$ for a polynomial $q = q(n)$.

- Can she predict $f_s(x^*)$ for some x^* of her choosing where $x^* \notin \{x_1,\ldots, x_q\}$? How well can she do it?

Lemma: If she succeeds with probability $\frac{1}{2^m} + 1/\mathrm{poly}(n)$, then she broke PRF security.

A Simple Lemma about Unpredictability

Let $f_s: \{0,1\}^\ell \to \{0,1\}^m$ be a pseudorandom function.

- Consider an adversary who requests and obtains $f_s(x_1), \dots, f_s(x_q)$ for a polynomial $q = q(n)$.

- Can she predict $f_s(x^*)$ for some x^* of her choosing where $x^* \notin \{x_1, \dots, x_q\}$? How well can she do it?

- Indistinguishability $\Longrightarrow$ Unpredictability (*but not vice versa*).

- Unpredictability $\equiv$ Indistinguishability *for bits* (lecture 3)

Challenge-Response Protocol

PRF Key s

Random r

$(ID, f_s(r))$

(ID number ID, PRF Key s)

"Proof": Adversary collects $(r_i, f_s(r_i))$ for poly many r_i (potentially of her choosing). She eventually has to produce $f_s(r^*)$ for a fresh random r^* when she is trying to impersonate.

This is hard as long as the input and output lengths of the PRF are long enough, i.e. $\omega(\log n)$.

TODAY

Applications of Pseudo-Random Functions (PRF):

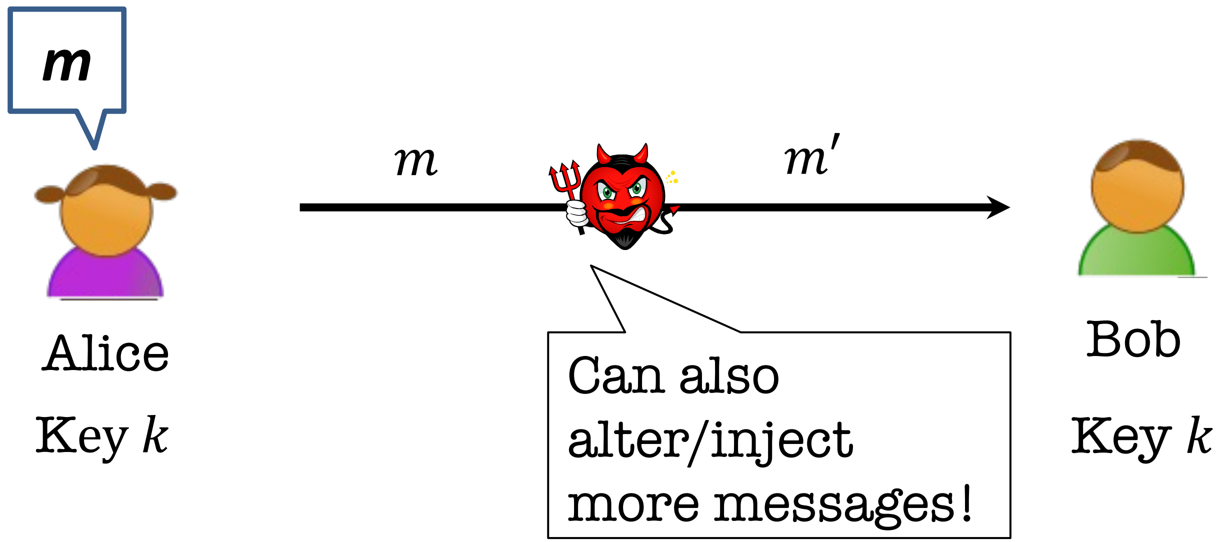
a. Identification Protocols ✓

b. Authentication

c. Encryption secure against Active Attacks

d. Applications to Learning Theory

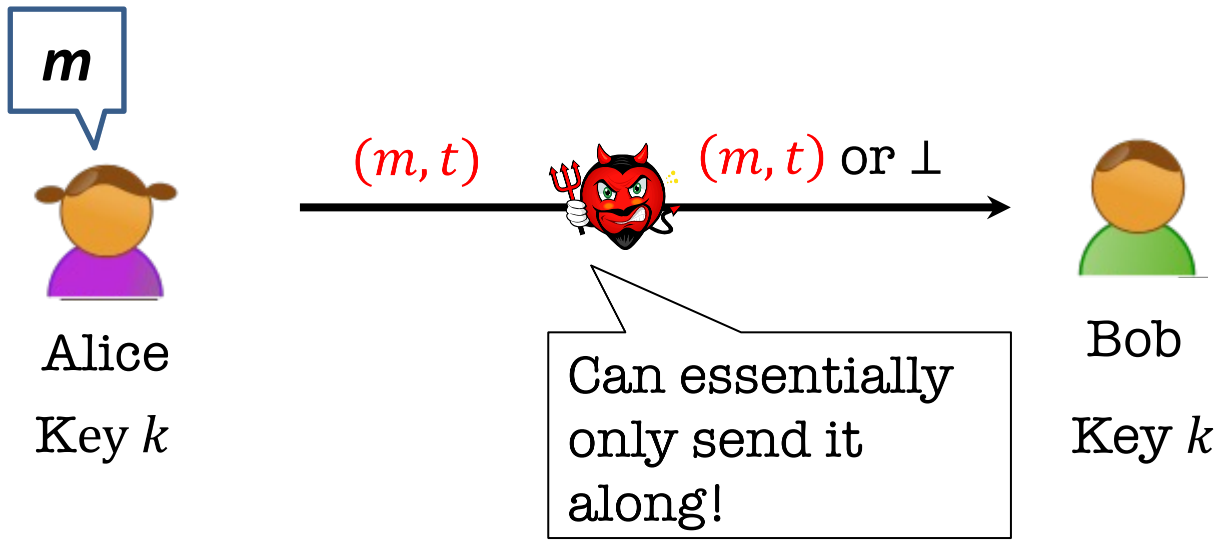
The authentication problem



This is known as a **man-in-the-middle attack.**

How can Bob check if the message is indeed from Alice?

The authentication problem



We want Alice to generate a tag for the message *m* which is **hard to generate** without the secret key *k*.

Message Authentication Codes (MACs)

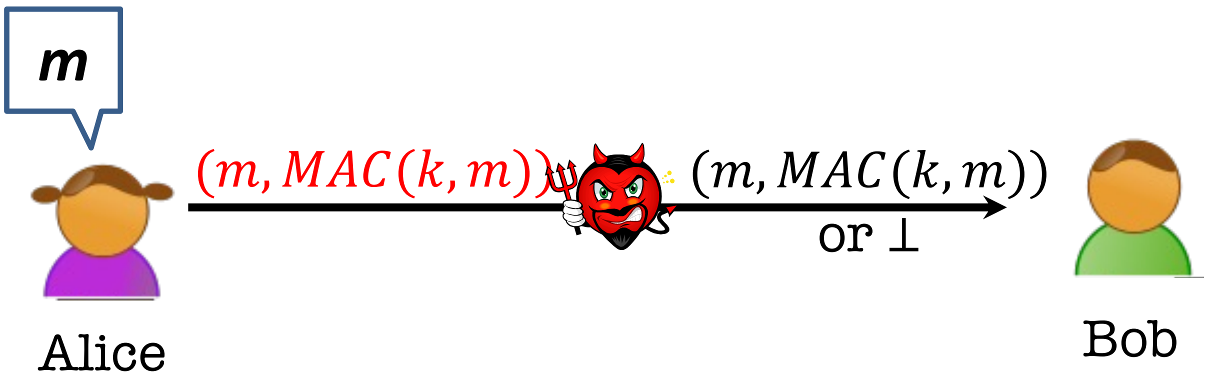
A triple of algorithms (Gen, MAC, Ver):

- Gen(1^n): Produces a key $k \leftarrow K$.
- MAC(k, m): Outputs a tag t (may be deterministic).
- Ver(k, m, t): Outputs Accept or Reject.

Correctness: Pr[Ver$\left(k, m, MAC(k, m)\right) =$ Accept] = 1

Security: *Hard to forge.* Intuitively, it should be hard to come up with a new pair *(m', t')* such that Ver accepts.

What is the power of the adversary?



- Can see many pairs $\left(m, MAC(k,m)\right)$.
- Can access a MAC oracle $MAC(k,\ \cdot\ )$
  - Obtain tags for message of choice.

This is called a *chosen message attack (CMA).*

Defining MAC Security

- **Total break:** The adversary should not be able to recover the key k.
- **Universal break:** The adversary can generate a valid tag for every message.
- **Existential break:** The adversary can generate a new valid tag t for some message m.

We will require MACs to be secure against the existential break!!

EUF-CMA Security

Existentially **U**n**f**orgeable against **C**hosen **M**essage **A**ttacks

m_1 →

← $t_1 = MAC(k, m_1)$

$k \leftarrow K$

m_2 →

← $t_2 = MAC(k, m_2)$

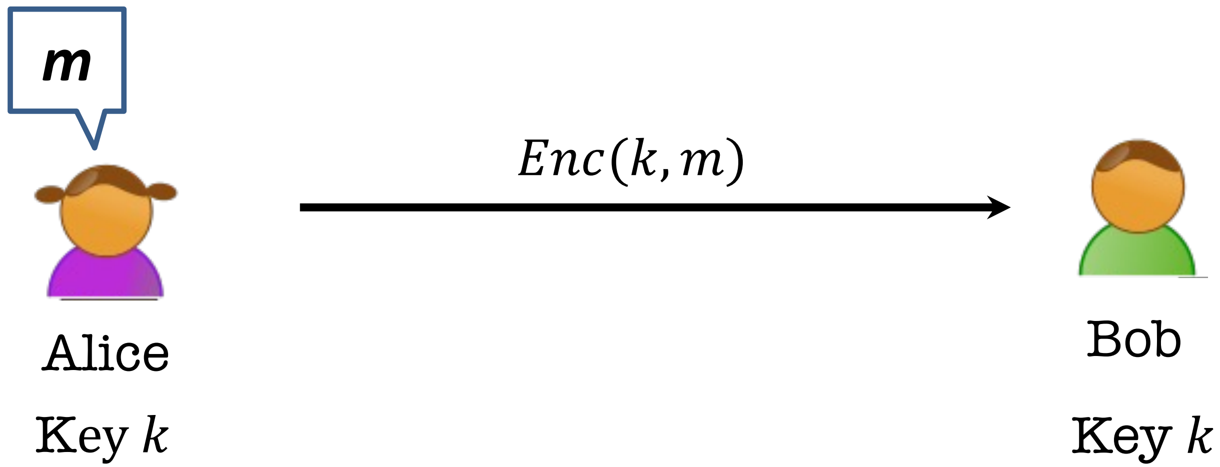
$\vdots$

(m, t) →

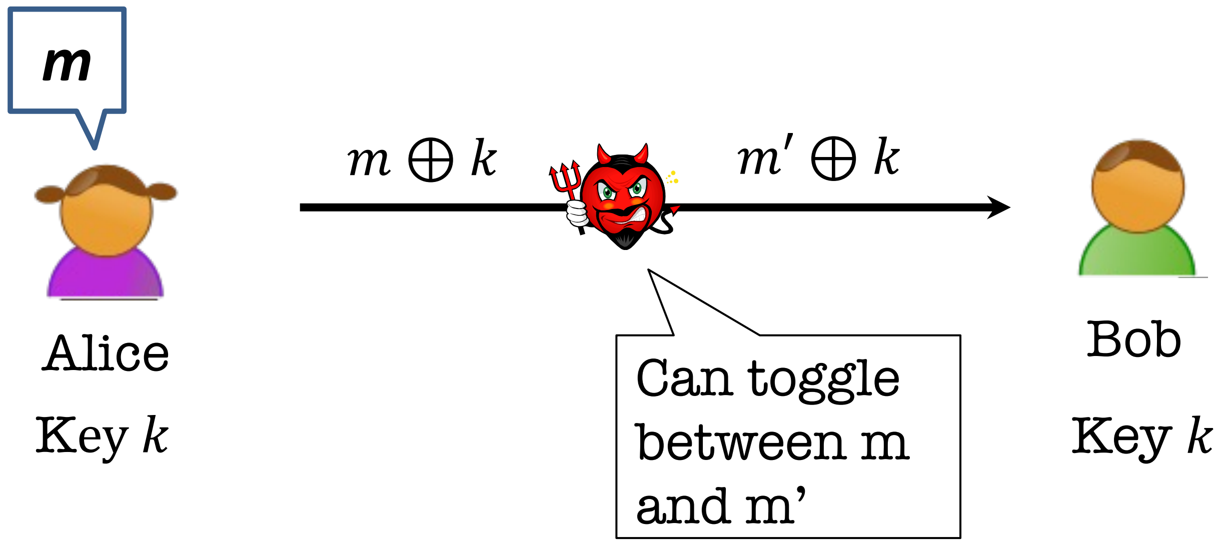
Accept if $(m, t) \neq (m_i, t_i)$ for all i, and $Ver(k, m, t) = 1$.

Want: $\Pr((m,t) \leftarrow A^{MAC(k,\cdot)}(1^n), Ver(k,m,t) = 1, (m,t) \notin Q)) = negl(n)$.
where Q is the set of queries $\{(m_i, t_i)\}_i$ that A makes.

Wait… Does encryption not solve this?

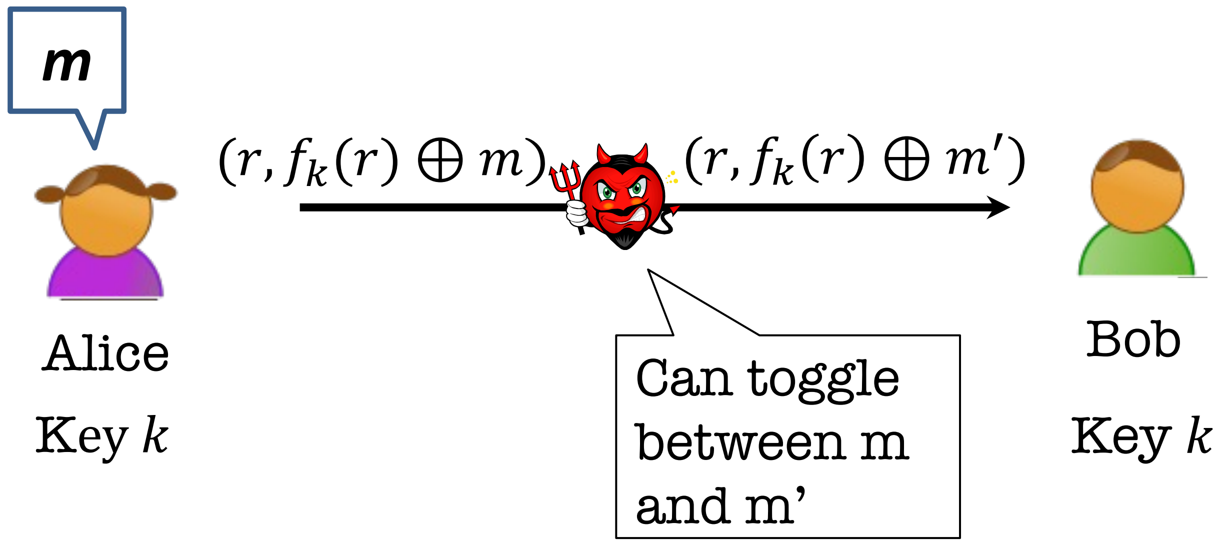


Wait… Does encryption not solve this?



One-time pad (and encryption schemes in general) are ***malleable***.

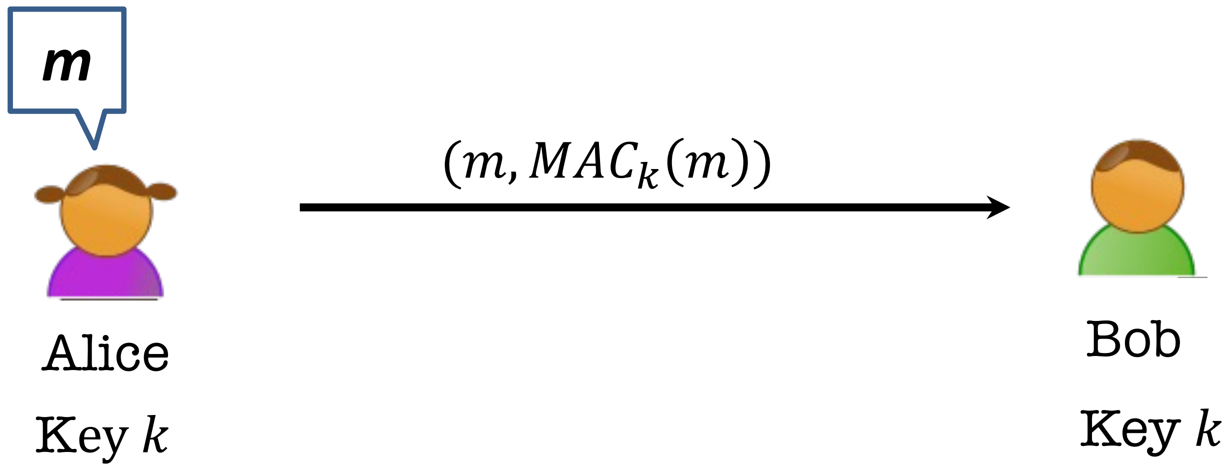
Wait… Does encryption not solve this?



One-time pad (and encryption schemes in general) are ***malleable***.

Privacy and Integrity are very **different goals!**

Constructing a MAC



Gen(1^n): Produces a PRF key $k \leftarrow K$.
MAC(k, m): Output $f_k(m)$.
Ver(k, m, t): Accept if $f_k(m) = t$, reject otherwise.

Security: Our earlier unpredictability lemma about PRFs essentially proves that this is secure!

Dealing with Replay Attacks

- The adversary could send an old valid *(m, tag)* at a later time.
  - In fact, our definition of security does not rule this out.

- **In practice:**
  - Append a time-stamp to the message. Eg. (m, T, MAC(m, T)) where T = 21 Sep 2022, 1:47pm.
  - Sequence numbers appended to the message (this requires the MAC algorithm to be *stateful*).

TODAY

Applications of Pseudo-Random Functions (PRF):

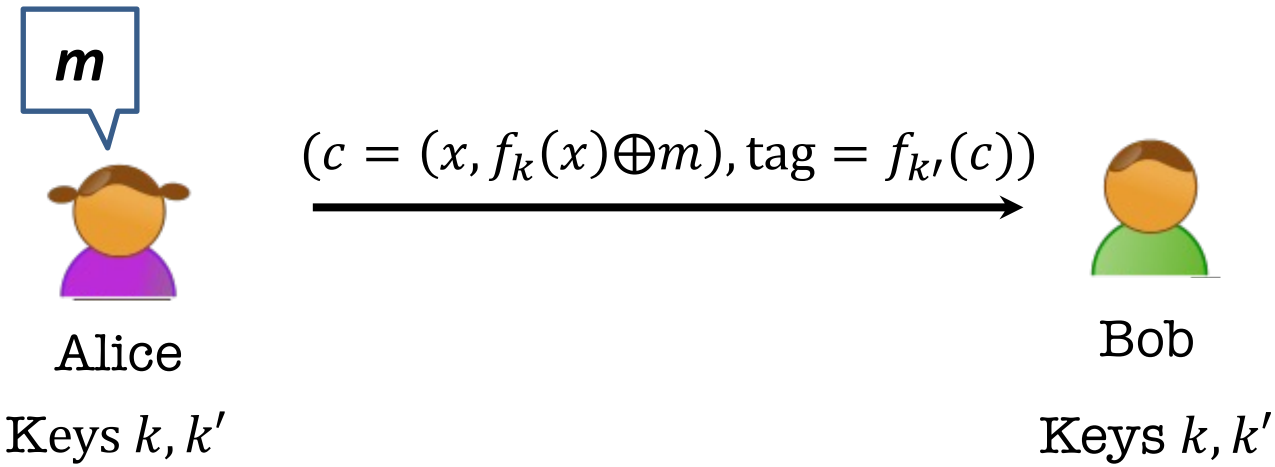
a. Identification Protocols ✓

b. Authentication ✓

c. Encryption secure against Active Attacks

d. Applications to Learning Theory

Privacy and Integrity!



MACs give us integrity, but not (necessarily) privacy.

Solution: Encrypt, then MAC!

TODAY

Applications of Pseudo-Random Functions (PRF):

a. Identification Protocols ✓

b. Authentication ✓

c. Encryption secure against Active Attacks ✓

d. Applications to Learning Theory

Negative Results in Learning Theory

Learning Theory / ML:

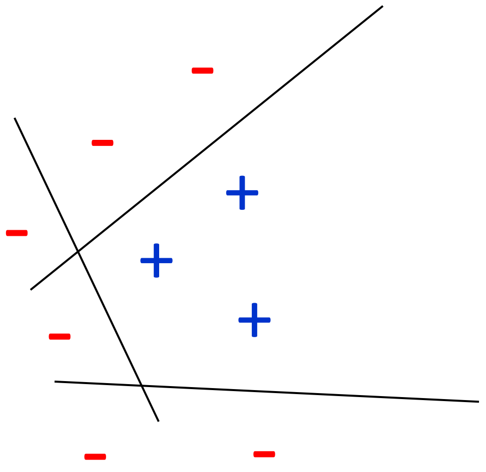
Given a few labeled examples $(x, f(x))$ for an unknown f, learn a hypothesis $h \approx f$

Cryptography (PRFs):

Construct function (families) $\{f\}$ for which it is hard to even predict f on a new input even given query-access to f.

Theorem [Kearns and Valiant 1994]:
Assuming PRFs exist, there are hypothesis classes that cannot be learned by polynomial-time algorithms.

Lots of More Negative Results…



Intersections of halfspaces

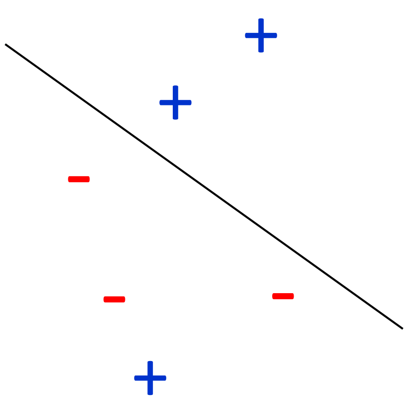
Cryptographic Hardness for Learning Intersections of Halfspaces

Adam R. Klivans *

University of Texas at Austin, Department of Computer Sciences, Austin, TX 78712 USA

Alexander A. Sherstov

University of Texas at Austin, Department of Computer Sciences, Austin, TX 78712 USA



“Agnostic learning” of halfspaces

Hardness of Agnostically Learning Halfspaces from Worst-Case Lattice Problems*

Stefan Tiegel†

February 21, 2023

Lots of More Applications…

Planting Undetectable Backdoors in Machine Learning Models

Shafi Goldwasser, UC Berkeley
Michael P. Kim, UC Berkeley
Vinod Vaikuntanathan, MIT
Or Zamir, IAS

On the Cryptographic Hardness of Learning Single Periodic Neurons

Min Jae Song*
Courant Institute
New York University
minjae.song@nyu.edu

Ilias Zadik*
Department of Mathematics
Massachusetts Institute of Technology
izadik@mit.edu

Joan Bruna
Courant Institute
Center for Data Science
New York University
bruna@cims.nyu.edu

Continuous LWE is as Hard as LWE & Applications to Learning Gaussian Mixtures

Aparna Gupte*
MIT
agupte@mit.edu

Neekon Vafa†
MIT
nvafa@mit.edu

Vinod Vaikuntanathan‡
MIT
vinodv@mit.edu

Watermarking GPT Outputs

Scott Aaronson (UT Austin and OpenAI)
Joint work with Hendrik Kirchner (OpenAI)

TODAY

Applications of Pseudo-Random Functions (PRF):

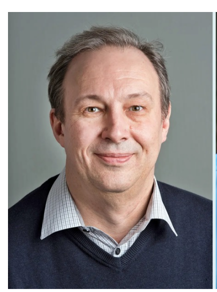
a. Identification Protocols ✓

b. Authentication ✓

c. Encryption secure against Active Attacks ✓

d. Applications to Learning Theory ✓

e. Applications to Complexity Theory: Natural Proofs



Razborov

Rudich