Understanding Information Disclosure from Secure Computation Output

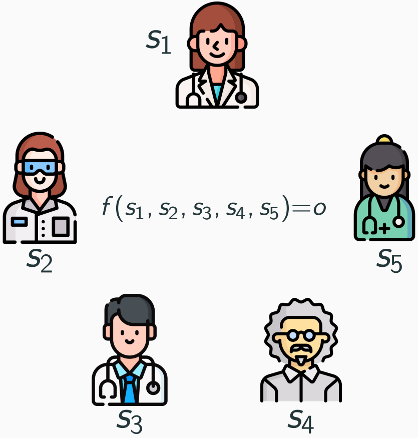
Analytical and Data-Driven Analysis

Alessandro Baccarini, PhD

abaccarini@proton.me
abaccarini.github.io

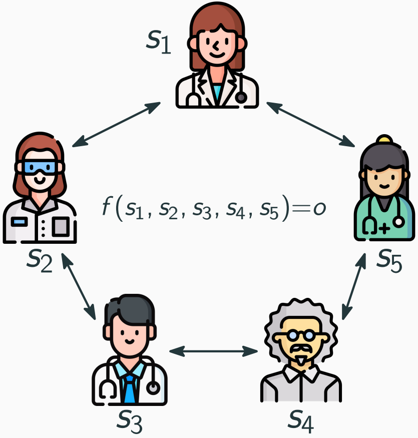
April 16, 2025

Motivation for secure computation



– Want to perform computation on private data

Motivation for secure computation



- Want to perform computation on private data
- Employ **cryptographic techniques** to compute without "seeing" the data

Last several decades of secure computation

- Emphasis has been on the **techniques**
- Efficiency, performance improvements
- Development tools to improve accessibility

- Emphasis has been on the **techniques**
- Efficiency, performance improvements
- Development tools to improve accessibility

Is this sufficient?

A glimpse into the not-too-distant past...

Technical Report

Achieving Higher-Fidelity Conjunction Analyses Using Cryptography to Improve Information Sharing

Brett Hemenway, William Welser IV, Dave Baiocchi

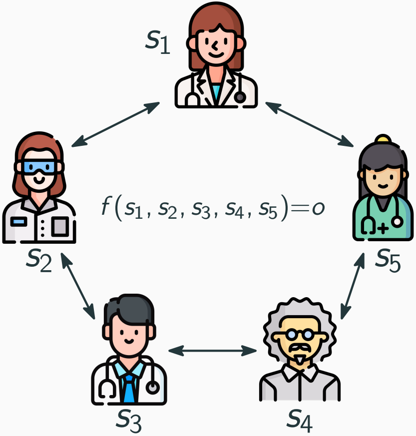
RAND CORPORATION

"There are many situations, however, where the output of the protocol itself may leak too much information...this leakage seems to be acceptable to the community, but this is a question that needs to be addressed before any MPC protocol can be securely deployed."

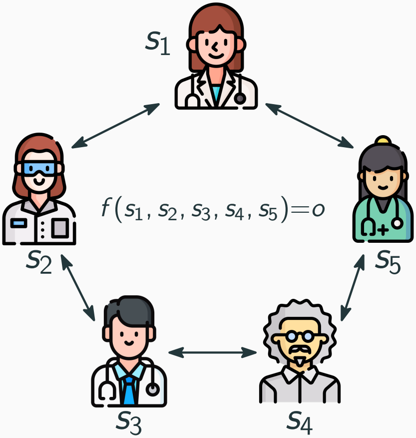
[HWB14]

A broader view of secure computation

Conventional security definitions distinguish between **unauthorized** and **authorized** leakage:



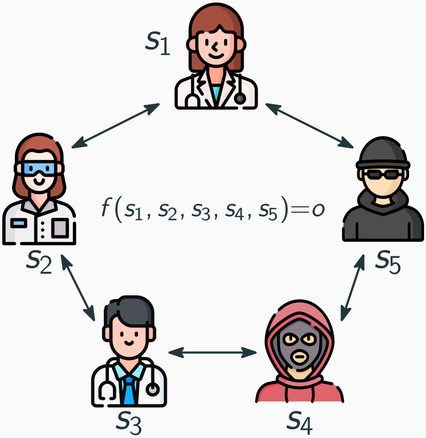
Conventional security definitions distinguish between **unauthorized** and **authorized** leakage:



Unauthorized leakage is eliminated (by design)

- Nothing is disclosed throughout the computation

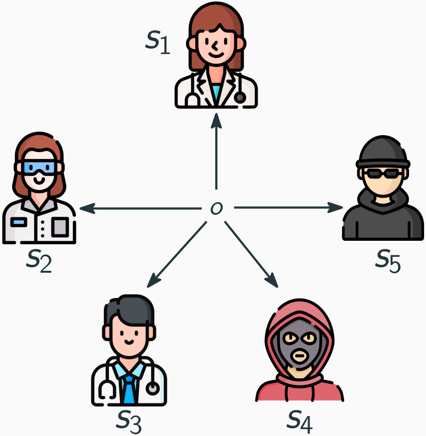
Conventional security definitions distinguish between **unauthorized** and **authorized** leakage:



Unauthorized leakage is eliminated (by design)

- Nothing is disclosed throughout the computation

Conventional security definitions distinguish between **unauthorized** and **authorized** leakage:



Unauthorized leakage is eliminated (by design)

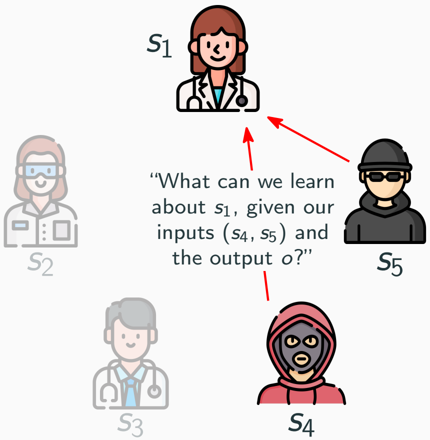
- Nothing is disclosed throughout the computation

Authorized (acceptable) leakage is **unconstrained**

- Does the **output** contain sensitive information?
- Can we **quantify** this leakage in a meaningful way?

A broader view of secure computation

Conventional security definitions distinguish between **unauthorized** and **authorized** leakage:



Unauthorized leakage is eliminated (by design)

- Nothing is disclosed throughout the computation

Authorized (acceptable) leakage is **unconstrained**

- Does the **output** contain sensitive information?
- Can we **quantify** this leakage in a meaningful way?

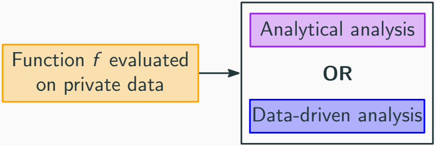
Information disclosure analysis [Bac24, Part II]

- Develop an information-theoretic approach to quantify leakage
- Apply technique to a practically significant function(s)
- Determine and apply appropriate mitigation strategies

Function f evaluated
on private data

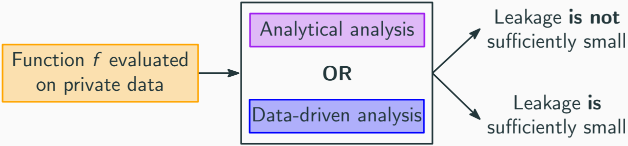
Information disclosure analysis [Bac24, Part II]

- Develop an information-theoretic approach to quantify leakage
- Apply technique to a practically significant function(s)
- Determine and apply appropriate mitigation strategies



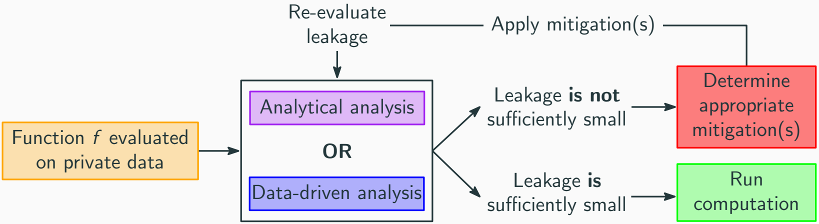
Information disclosure analysis [Bac24, Part II]

- Develop an information-theoretic approach to quantify leakage
- Apply technique to a practically significant function(s)
- Determine and apply appropriate mitigation strategies

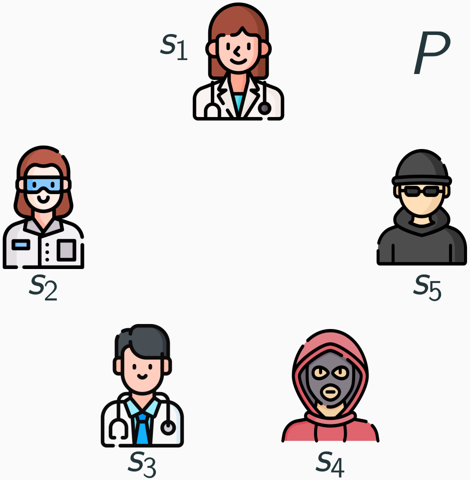


Information disclosure analysis [Bac24, Part II]

- Develop an information-theoretic approach to quantify leakage
- Apply technique to a practically significant function(s)
- Determine and apply appropriate mitigation strategies

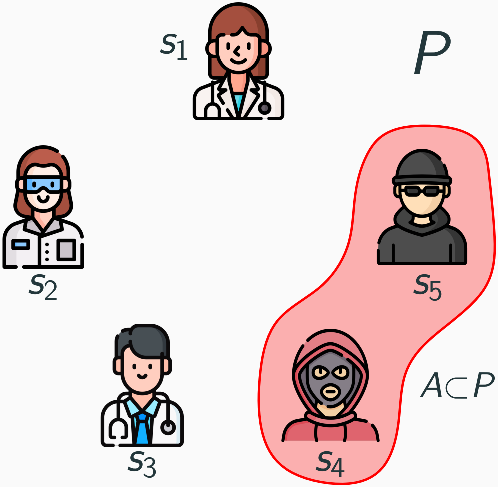


Establishing the setting



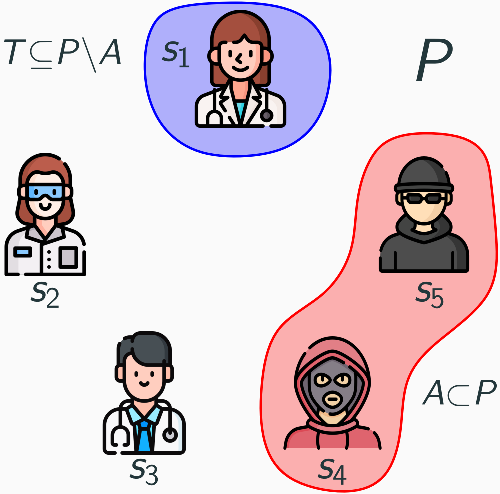
Partition parties P into:

Establishing the setting



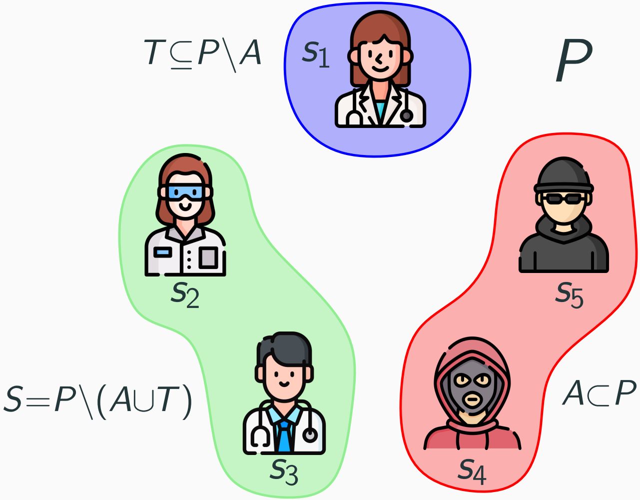
Partition parties P into: attackers A,

Establishing the setting



Partition parties P into: attackers A, targets T,

Establishing the setting



Partition parties P into: attackers A, targets T, spectators S

Metric?

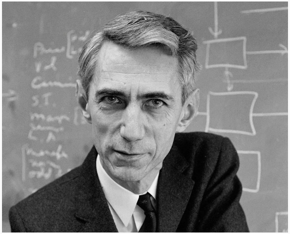
Model participant inputs by a **random variable** X_{P_i}

How to **measure** the **information** disclosed by the output?

Model participant inputs by a **random variable** X_{P_i}

How to **measure** the **information** disclosed by the output?

Entropy!



C. Shannon. Photo: Alfred Eisenstaedt

Shannon
$H(X)$ (discrete)

Differential
$h(X)$ (continuous)

Putting everything together

Attackers $\mathbf{X}_A$, targets $\mathbf{X}_T$, and spectators $\mathbf{X}_S$ (vectors)

Treat the **output** as a random variable: $f(\mathbf{X}_A, \mathbf{X}_T, \mathbf{X}_S) = O$

Putting everything together

Attackers $\mathbf{X}_A$, targets $\mathbf{X}_T$, and spectators $\mathbf{X}_S$ (vectors)

Treat the **output** as a random variable: $f(\mathbf{X}_A, \mathbf{X}_T, \mathbf{X}_S) = O$

Attackers' weighted average entropy **[AH17]**

$H(\mathbf{X}_T \mid \mathbf{X}_A = \mathbf{x}_A, O)$ how much information A learns about T, given $\mathbf{x}_A$ and O

Absolute entropy loss **[BBZ24a; BBZ24b]**

$H(\mathbf{X}_T) - H(\mathbf{X}_T \mid \mathbf{X}_A = \mathbf{x}_A, O)$ the **total** amount of information disclosed about T, given $\mathbf{x}_A$ and O

Absolute entropy loss $\iff$ **mutual information** between $\mathbf{X}_T$ and O (conditioned on $\mathbf{X}_A$)

Case study: the average salary computation

- 2016 Boston gender pay gap survey
- Analyzed the **private** wages based on gender and race using **multi-party computation**
- **Average (salary) computation**

Mayor Walsh & Boston Women's Workforce Council Release 2016 Gender Wage Gap Report; New Partnership with BU

Thursday, January 5, 2017 – Mayor Martin J. Walsh and the Boston Women's Workforce Council (BWWC) released the 2016 gender wage report and announced a

Source: Boston University, 2017

Case study: the average salary computation

- 2016 Boston gender pay gap survey
- Analyzed the **private** wages based on gender and race using **multi-party computation**
- **Average (salary) computation**

Mayor Walsh & Boston Women's Workforce Council Release 2016 Gender Wage Gap Report; New Partnership with BU

Thursday, January 5, 2017 – Mayor Martin J. Walsh and the Boston Women's Workforce Council (BWWC) released the 2016 gender wage report and announced a

Source: Boston University, 2017

However, the average reduces to a **sum**:

$$f_\mu(\mathbf{x}) = \frac{1}{n}(x_1 + \cdots + x_n) \longrightarrow x_1 + \cdots + x_n$$

Case study: the average salary computation

- 2016 Boston gender pay gap survey
- Analyzed the **private** wages based on gender and race using **multi-party computation**
- **Average (salary) computation**

Mayor Walsh & Boston Women's Workforce Council Release 2016 Gender Wage Gap Report; New Partnership with BU

Thursday, January 5, 2017 – Mayor Martin J. Walsh and the Boston Women's Workforce Council (BWWC) released the 2016 gender wage report and announced a

Source: Boston University, 2017

However, the average reduces to a **sum**:

$f_\mu(\mathbf{x}) = \frac{1}{n}(x_1 + \cdots + x_n) \longrightarrow x_1 + \cdots + x_n$

$$\boxed{O = \sum_i X_{T_i} + \sum_j X_{A_j} + \sum_k X_{S_k}}$$

Optimal attacker strategy

Are there certain input(s) an attacker can supply to **maximize** the information they learn?

Optimal attacker strategy

Are there certain input(s) an attacker can supply to **maximize** the information they learn?

Claim

The information disclosure is **independent** of the attackers' input(s).

Optimal attacker strategy

Are there certain input(s) an attacker can supply to **maximize** the information they learn?

Claim

The information disclosure is **independent** of the attackers' input(s).

– Intuition: an adversary can "remove" their influence

s_5 

s_4 

$$o = \cancel{\tfrac{1}{5}}(s_1 + s_2 + s_3 + s_4 + s_5)$$

Are there certain input(s) an attacker can supply to **maximize** the information they learn?

Claim

The information disclosure is **independent** of the attackers' input(s).

– Intuition: an adversary can "remove" their influence

s_5

s_4

$o = s_1 + s_2 + s_3 + s_4 + s_5$

Optimal attacker strategy

Are there certain input(s) an attacker can supply to **maximize** the information they learn?

Claim

The information disclosure is **independent** of the attackers' input(s).

– Intuition: an adversary can "remove" their influence

s_5

s_4

$$o = s_1 + s_2 + s_3 + s_4 + s_5$$

Optimal attacker strategy

Are there certain input(s) an attacker can supply to **maximize** the information they learn?

Claim

The information disclosure is **independent** of the attackers' input(s).

– Intuition: an adversary can "remove" their influence

s_5

s_4

$$o = s_1 + s_2 + s_3$$

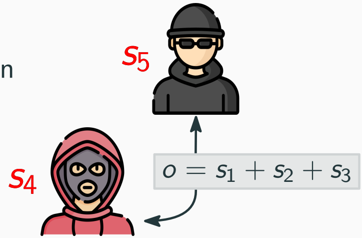
Optimal attacker strategy

Are there certain input(s) an attacker can supply to **maximize** the information they learn?

Claim

The information disclosure is **independent** of the attackers' input(s).

- Intuition: an adversary can "remove" their influence
- Not universally true (depends on f)



Single evaluation

- Model inputs by common distributions:
  - Poisson
  - Uniform
  - Gaussian
  - Log-normal [Cao+22]

Single evaluation

- Model inputs by common distributions:
    - Poisson
    - Uniform
    - **Gaussian**
    - Log-normal [Cao+22]
- For a single evaluation, information disclosure is **independent** of
    - the distribution parameters
    - the **distribution itself**
- Disclosure is proportional to the number of **spectators**

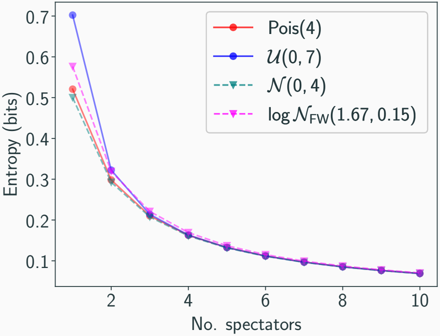


Figure 1: Absolute entropy loss (lower is better), 1 target

Computing the average salary *twice*

- First study was a success
- Repeated the following year with an extended set of participants

Mayor Walsh & BWWC Release 2017 Wage Gap Report

The Boston Women's Workforce Council released its 2017

BOSTON WOMEN'S WORKFORCE COUNCIL REPORT 2017

Source: Boston University, 2018

Computing the average salary *twice*

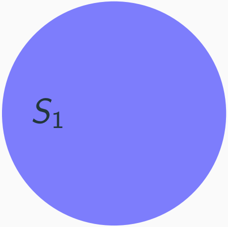
- First study was a success
- Repeated the following year with an extended set of participants
- Spectators present in the first, second, and both evaluation(s)

Mayor Walsh & BWWC Release 2017 Wage Gap Report

The Boston Women's Workforce Council released its 2017

BOSTON WOMEN'S WORKFORCE COUNCIL REPORT 2017

Source: Boston University, 2018



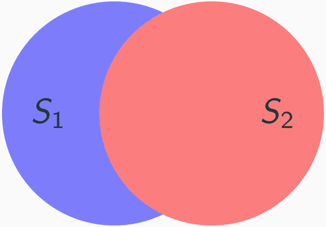
Computing the average salary *twice*

- First study was a success
- Repeated the following year with an extended set of participants
- Spectators present in the first, second, and both evaluation(s)

Mayor Walsh & BWWC Release 2017 Wage Gap Report

The Boston Women's Workforce Council released its 2017

Source: Boston University, 2018



Computing the average salary *twice*

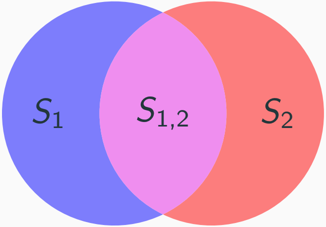
- First study was a success
- Repeated the following year with an extended set of participants
- Spectators present in the first, second, and both evaluation(s)
  - **Correlated** outputs

Mayor Walsh & BWWC Release 2017 Wage Gap Report

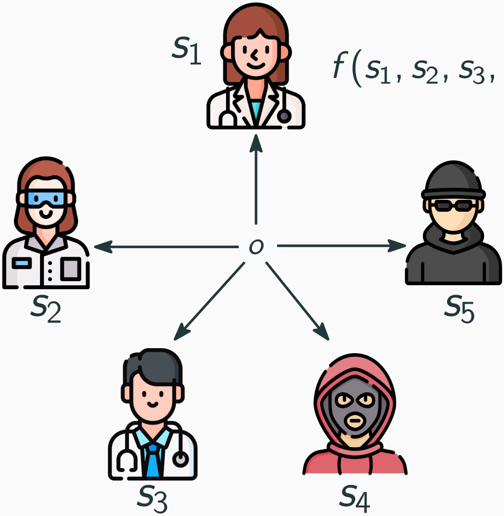
The Boston Women's Workforce Council released its 2017

BOSTON WOMEN'S WORKFORCE COUNCIL REPORT 2017

Source: Boston University, 2018



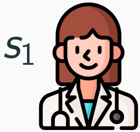
An interesting question



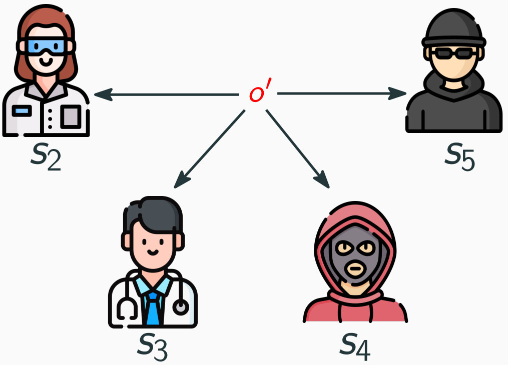
$$f(s_1, s_2, s_3, s_4, s_5) = o$$

An interesting question

"What happens if everyone else participates *again*, but without me?"



$$f(s_2, s_3, s_4, s_5) = o'$$



Target participates in one or both evaluations

Vary the **ratio** of shared spectators $S_{1,2}$ to the (fixed) total number of spectators

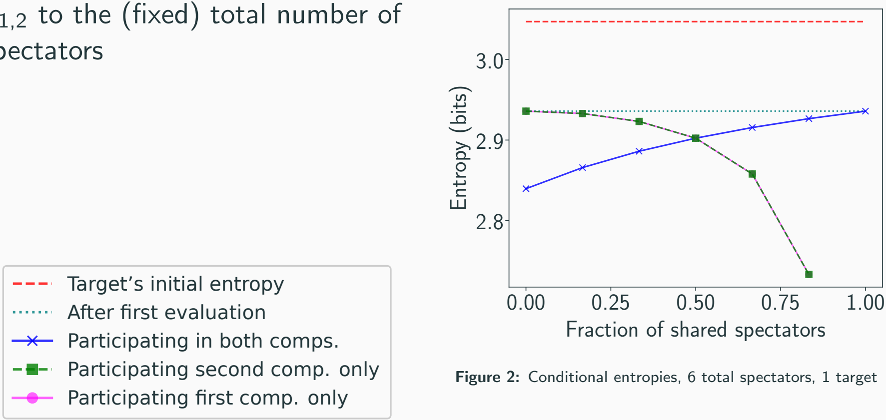
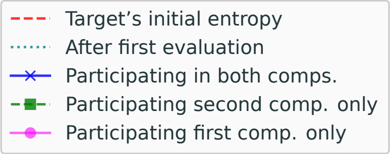


Figure 2: Conditional entropies, 6 total spectators, 1 target

Target participates in one or both evaluations

Vary the **ratio** of shared spectators $S_{1,2}$ to the (fixed) total number of spectators

- Largest protection at **50% overlap**
- Undesirable disclosure at extrema



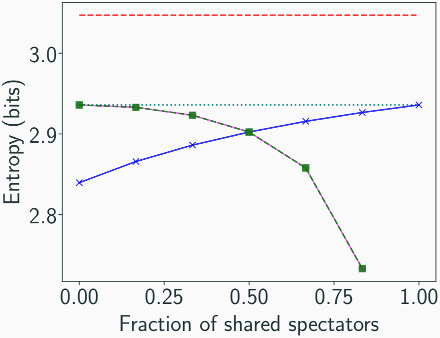


Figure 2: Conditional entropies, 6 total spectators, 1 target

Next step: advanced statistical measures

What are some logical successors to the average?

Next step: advanced statistical measures

What are some logical successors to the average?

- Order statistics (max/min, median)

$$f_{\max}(\mathbf{x}) = \max_i x_i$$

- Variability measures (variance)

$$f_{\sigma^2}(\mathbf{x}) = \frac{1}{n} \sum_i (x_i - f_\mu(\mathbf{x}))^2$$

- *Multidimensional* functions

$$f_{(\mu,\sigma^2)}(\mathbf{x}) = (f_\mu(\mathbf{x}), f_{\sigma^2}(\mathbf{x}))$$

New functions $\implies$ new challenges

Prior analysis leveraged properties of **sums of RVs**, **closed-form expressions** of the entropy

New functions $\implies$ new challenges

Prior analysis leveraged properties of **sums of RVs**, **closed-form expressions** of the entropy

Data-driven entropy estimators

Discrete
plug-in

Continuous
k-nearest neighbors

New functions $\implies$ new challenges

Prior analysis leveraged properties of **sums of RVs**, **closed-form expressions** of the entropy

Data-driven entropy estimators

Discrete
plug-in

Continuous
k-nearest neighbors

Problem

Function could produce discrete outputs from continuous inputs, producing a "**mixture**"

New functions $\implies$ new challenges

Prior analysis leveraged properties of **sums of RVs**, **closed-form expressions** of the entropy

Data-driven entropy estimators

Discrete
plug-in

Continuous
k-nearest neighbors

Problem

Function could produce discrete outputs from continuous inputs, producing a "**mixture**"

Computer Science > Information Theory

[Submitted on 19 Sep 2017 (v1), last revised 9 Oct 2018 (this version, v3)]

Estimating Mutual Information for Discrete-Continuous Mixtures

Weihao Gao, Sreeram Kannan, Sewoong Oh, Pramod Viswanath

New functions $\implies$ new challenges

Prior analysis leveraged properties of **sums of RVs**, **closed-form expressions** of the entropy

Data-driven entropy estimators

Discrete
plug-in

Continuous
k-nearest neighbors

Problem

Function could produce discrete outputs from continuous inputs, producing a "**mixture**"

Computer Science > Information Theory

[Submitted on 19 Sep 2017 (v1), last revised 9 Oct 2018 (this version, v3)]

Estimating Mutual Information for Discrete-Continuous Mixtures

Weihao Gao, Sreeram Kannan, Sewoong Oh, Pramod Viswanath

Recall (from slide 9). . .

mutual information

$\Leftrightarrow$

absolute loss

Intuitive observations: maximum

Maximum

An adversary **maximizes** the information learned by **minimizing** their influence.*

*Inverse behavior for the minimum

Intuitive observations: maximum

Maximum

An adversary **maximizes** the information learned by **minimizing** their influence.*

*Inverse behavior for the minimum

In fact, the information A learns is **bounded** by observing the output, **without participating** in $f_{\max}$

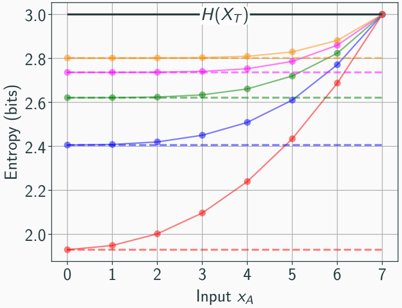


Figure 3: Uniform $\mathcal{U}(0, 7)$, $H(\mathbf{X}_T | \mathbf{X}_A = \mathbf{x}_A, O)$, 1 target

Intuitive observations: maximum

Maximum

An adversary **maximizes** the information learned by **minimizing** their influence.*

*Inverse behavior for the minimum

In fact, the information A learns is **bounded** by observing the output, **without participating** in $f_{\max}$

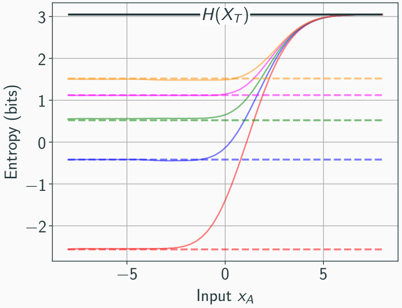


Figure 3: Normal $\mathcal{N}(0, 4)$, $H(\mathbf{X_T}|\mathbf{X}_A = \mathbf{x}_A, O)$, 1 target

Intuitive observations: maximum

Maximum

An adversary **maximizes** the information learned by **minimizing** their influence.*

*Inverse behavior for the minimum

In fact, the information A learns is **bounded** by observing the output, **without participating** in $f_{\max}$

– Proof is a work-in-progress

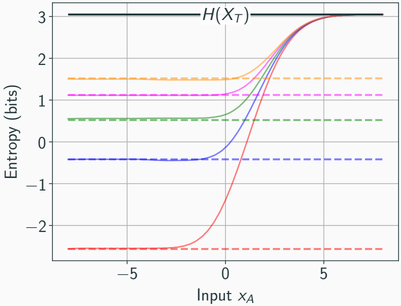


Figure 3: Normal $\mathcal{N}(0,4)$, $H(\mathbf{X}_T|\mathbf{X}_A = \mathbf{x}_A, O)$, 1 target

Ongoing work

Data-driven analysis serves as a "first pass" assessment of a function's suitability for secure computation

Ongoing work

Data-driven analysis serves as a "first pass" assessment of a function's suitability for secure computation

Further analysis of complex funcs.

- Derive analytical expressions information disclosure
- Apply data-driven analysis to broader functionalities

Ongoing work

Data-driven analysis serves as a "first pass" assessment of a function's suitability for secure computation

Further analysis of complex funcs.

- Derive analytical expressions information disclosure
- Apply data-driven analysis to broader functionalities

Mitigation strategies

- Adding noise (differential privacy)
- Introducing synthetic inputs
- Modifying the function

Conclusions and takeaways

- Despite decades of performance improvements, broader privacy concerns remain that must be addressed prior to deployment of secure computation
- Developed a framework for quantifying information disclosure from secure computation outputs
- Computation designers can use this framework to determine potential disclosure about participants' inputs

Thank you!

Questions?

References

[AH17] P. Ah-Fat and M. Huth. "Secure Multi-party Computation: Information Flow of Outputs and Game Theory". In: *International Conference on Principles of Security and Trust*. 2017, pp. 71–92.

[Bac24] A. Baccarini. "New Directions in Secure Multi-Party Computation: Techniques and Information Disclosure Analysis". PhD Thesis. University at Buffalo, 2024.

[BBZ24a] A. Baccarini, M. Blanton, and S. Zou. "Understanding Information Disclosure from Secure Computation Output: A Comprehensive Study of Average Salary Computation". In: *ACM Transactions on Privacy and Security (TOPS)* 28.1 (2024), pp. 1–36.

[BBZ24b] A. Baccarini, M. Blanton, and S. Zou. "Understanding Information Disclosure from Secure Computation Output: A Study of Average Salary Computation". In: *ACM CODASPY*. 2024, pp. 187–198.

[Cao+22] L. Cao, T. Tong, D. Trafimow, T. Wang, and X. Chen. "The A Priori Procedure for estimating the mean in both log-normal and gamma populations and robustness for assumption violations". In: *Methodology* 18.1 (2022), pp. 24–43.

[Gao+17] W. Gao, S. Kannan, S. Oh, and P. Viswanath. "Estimating mutual information for discrete-continuous mixtures". In: *Proceedings on Advances in Neural Information Processing Systems (NeurIPS)* 30 (2017), pp. 5988–5999.

[HWB14] B. Hemenway, W. Welser, and D. Baiocchi. *Achieving Higher-fidelity Conjunction Analyses Using Cryptography to Improve Information Sharing*. Tech. rep. RR-344-AF. Rand Corporation, Feb. 2014.