Privacy and Security Issues in Speech Processing

Bhiksha Raj

ATSIP, 24 May 2022

Acknowledgements..

- Thanks (in alphabetic order) to:
  - Abelino Jimenez, LinkedIn
  - Dijana Petrovska, Telecom Paris
  - Francisco Teixeira, INESC Pt
  - Gerard Chollet, Telecom Paris
  - Isabel Trancoso, INESC Pt
  - Jose Portelo, INESC Pt
  - Manas Pathak, Google
  - Paris Smaragdis, UIUC
  - Raphael Olivier, CMU
  - Shantanu Rane, PARC
  - Sohail Bahmani, GaTech
- And many other students, colleagues and friends who have all contributed to the work presented here

Introducing me

- Bhiksha Raj
  - Professor
  - Carnegie Mellon University
    - Language Technologies Inst.
    - Additional affiliations:
      - Electrical and Computer Engg.
      - Machine Learning
      - Music Technologies

- Primary areas of research:
  - Speech and audio processing
    - ASR systems
    - Algorithms, robustness to environmental conditions, low-resource conditions
    - Speech as biometric, and speech forensics
    - ***Trustworthiness, privacy and security of speech systems***

Speech technologies are not currently trustworthy!!

- Trustworthiness: accuracy, reliability, fairness, robustness, security…
    - The accuracy is often questionable
    - They break down under acoustic or linguistic variations/noise
    - Biased: Accent bias, gender bias, economically biased
        - Possibly among the most biased, yet popular technologies
    - They can be hacked by adversaries

- And they are challenged in terms of privacy!!

Speech technologies are not currently trustworthy!!

- Trustworthiness: accuracy, reliability, fairness, robustness, security…
  - The accuracy is often questionable
  - They break down under acoustic or linguistic variations/noise
  - Biased: Accent bias, gender bias, economically biased
    - Possibly among the most biased, yet popular technologies
  - They can be hacked by adversaries

- And they are challenged in terms of privacy!!
  - The topic of this talk

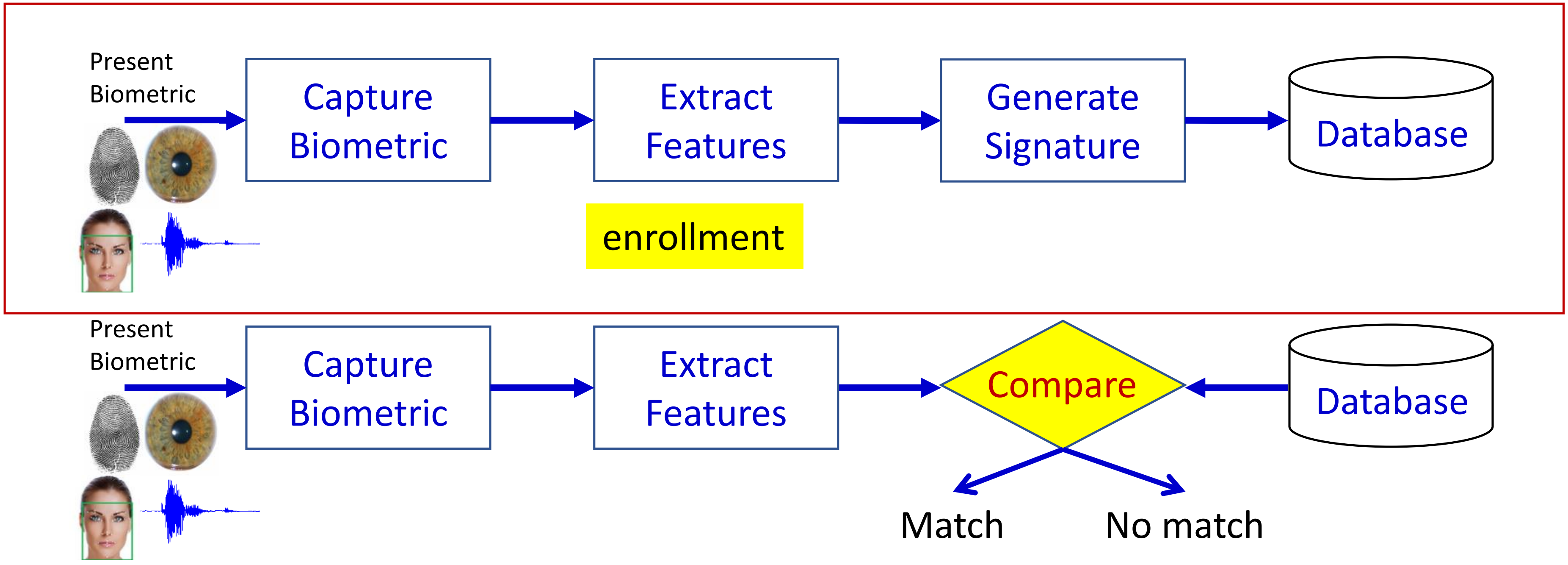
Speech is a highly informative signal

- Speech carries a lot more information than the mere content of what was spoken
  - Your identity

Speech is a biometric

Sufficiently distinctive to be usable for identity authentication

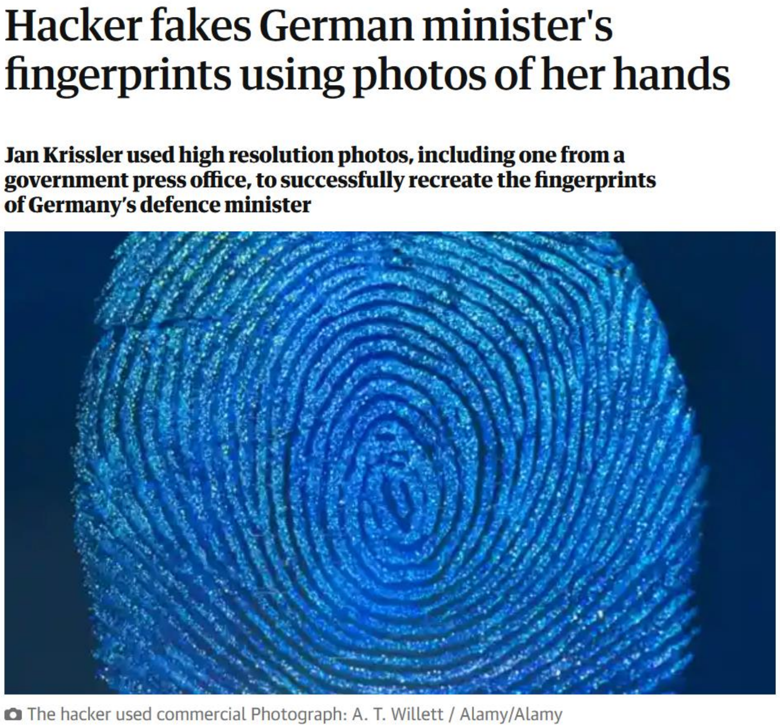
Biometric Challenges



- Typical biometric verification procedure

Laden with risks

As a biometric



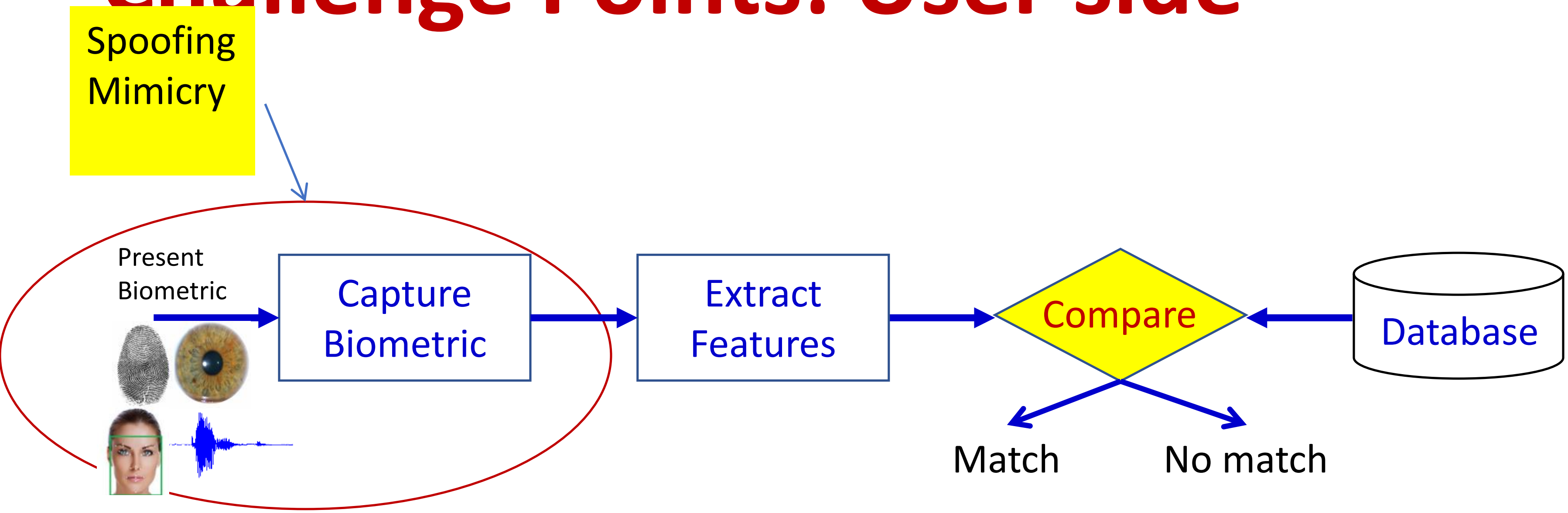
Hacker fakes German minister's fingerprints using photos of her hands

Jan Krissler used high resolution photos, including one from a government press office, to successfully recreate the fingerprints of Germany's defence minister

The hacker used commercial Photograph: A. T. Willett / Alamy/Alamy

- Unlike passwords, not cancellable…

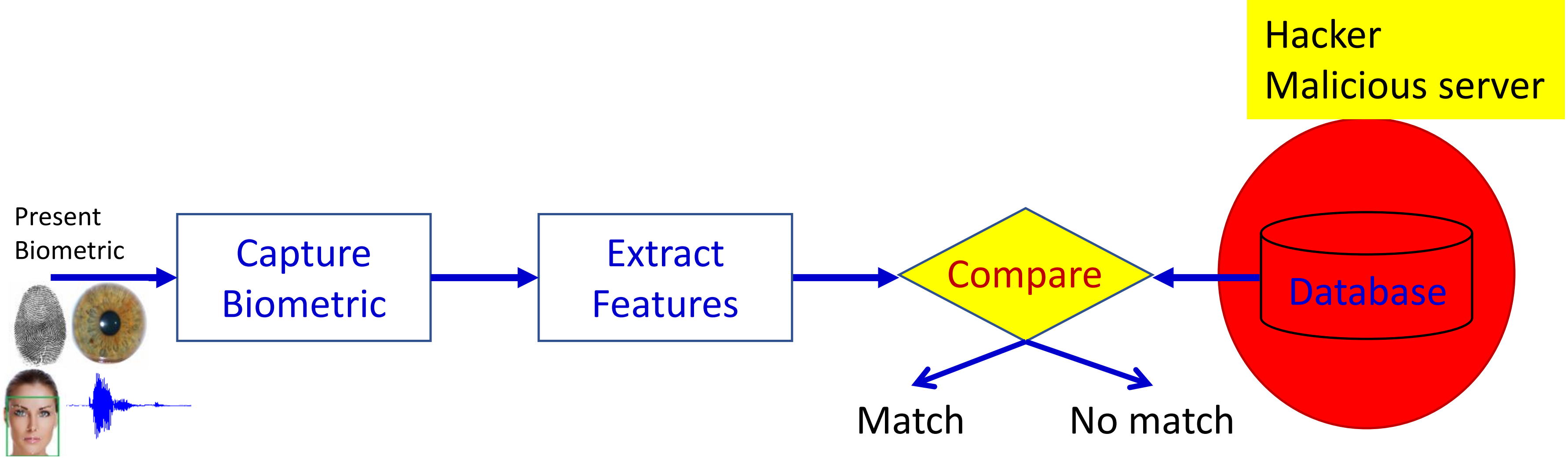
Challenge Points: User side



- **User side:** A spoofer or mimic can duplicate biometric
  - And can continue to compromise you for the rest of your life

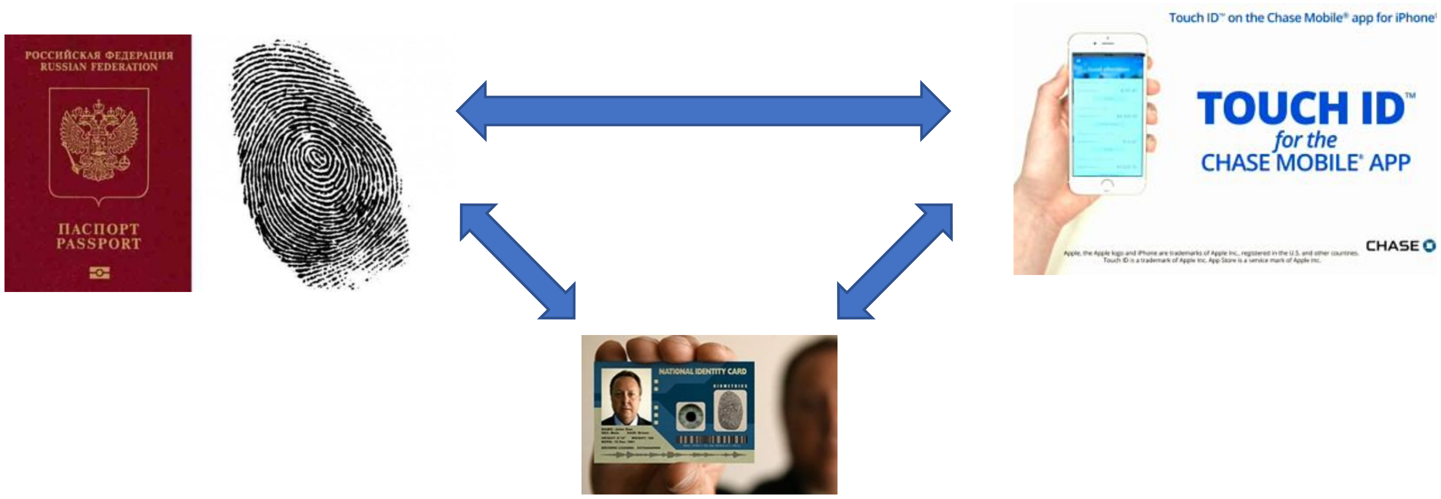
- **Biometrics are public:** Adversary could easily obtain biometrics from public venues

Challenge Points: System Side



- **System side:** Hacker or a malicious server has your biometric
  - Can use it to authenticate as you in other services/apps
  - Can use it to create synthetic data to mimic/spoof you
  - Can use it to track you in other places
    - E.g. use voice/IRIS/fingerprint/face to find/track you on YouTube and public fora.

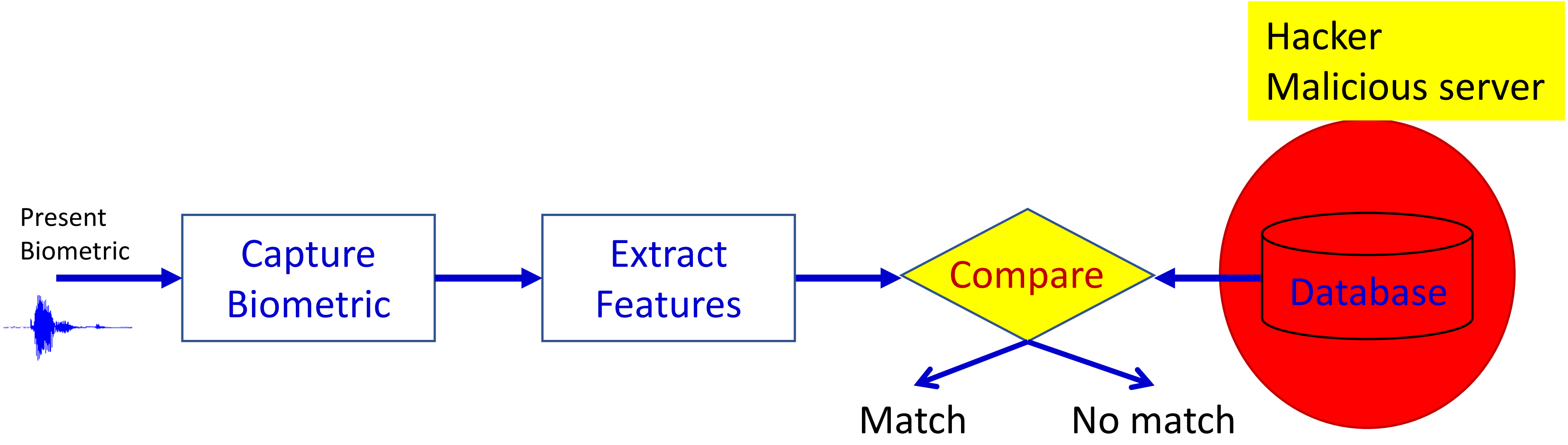
Challenge: Linkability



- Biometric signatures stored by different accounts are very similar
  - The fingerprint you store with the passport office is the same as your fingerprint at the bank is the same as your fingerprint in your biometric ID card

- They can be linked
  - Agencies could collude to track you
  - Big brother can watch you more carefully

Challenge Points: System Side



- **System side:** Hacker/Malicious server has biometric
  - Can learn undesired information about you
    - Voice carries information about health, ethnicity, gender, education, age, etc

The world recognizes the problem

- EU: General Data Protection Regulation (GDPR), 2016
    - Data controllers must adopt *technical and organisational measures* to implement data protection principles, such as data minimisation.
        - Does not *specifically* mention speech, but is implied

- Illinois: Biometric Information Privacy Act, 2008
    - No private entity may collect, capture, purchase, receive through trade, or otherwise obtain a person's or a customer's biometric identifier or biometric information.
        - Voiceprints included in definition
- Texas: Capture or Use of Biometric Identifier (CUBI) law, 2009
    - Regulates collection of (among other identifiers) voiceprints
- Washington: House Bill 1493, 2017
    - Places restrictions on the collection of (among other identifiers) voiceprints

- California Consumer Privacy Act (CCPA), 2018.
    - *Biometric information* is defined comprehensively to include not only physical characteristics but also behavioral ones. Expressly refers to voice recordings

ISO/IEC JTC 1/SC 27

- “ISO/IEC JTC 1/SC 27 Information security, cybersecurity and privacy protection”
  - Standardization subcommittee of the Joint Technical Committee ISO/IEC JTC 1 of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), 1989.

- Requirements for biometrics:
  - **Unlinkability:** not be possible to say whether two protected biometric sample representations belong to the same subject. This prevents cross-comparisons for databases of different applications and ensures the privacy of the subject.
  - **Renewability:** If a protected biometric reference is leaked or lost, the reference data can be revoked and renewed from the same biometric trait without the need to re-enroll.
  - **Irreversibility:** Recovering biometric data from leaked protected biometric information is impossible without knowing the secret used to protect the biometric information. Restoring of valid biometric features or samples is prevented

Speech is a highly informative signal

- Speech carries a lot more information than the mere content of what was spoken
  - Your identity

Speech is a biometric

Speech is a highly informative signal

- Speech carries a lot more information than the mere content of what was spoken
  - Your identity
  - Your gender
  - Your nationality
  - Your emotional status
  - Your health
  - Your educational status and other demographic information

Speech is a biometric

Speech is more than a biometric

Speech technologies could invade your privacy

- Every time you use SIRI to access a service you:
  - Give it a biometric signature of yourself
  - Inform it of your gender, your nationality, your emotional status, your health, etc. etc. etc.

- This is often *unintended* exposure and can be abused
  - E.g. a bank analyses your voice to determine that you are unhealthy and raises your interest rate
  - You can be traced, tracked, categorized, face biases or be discriminated for/against, etc.
  - Your private information risks exposure
  - Risk is amplified when the system is *always listening*

So, what are the challenges

- Usage: You want to use the system but are concerned that
  - Your biometric could be compromised
    - Possibly by the service provider itself
    - Possibly permanently
  - Your service could make undesired inferences about you

- Sharing: You may be willing to volunteer to contribute *your* voice to improve models
  - But are concerned about what the system/outside user/hacker could learn from it

Requirement: Usage

- **Usage/inference:** Somehow process the data such that:
  - **Biometrics:**
    - The system cannot link to other sources of voice
    - The biometric is revocable if lost
    - The system cannot reverse engineer biometric to know more about the speaker
    - The “system” cannot make any inferences besides what is permitted
      - E.g. recognize speech, but not learn about the speaker’s ID/gender/other info
      - E.g. biometrically verify the speaker, but be unable to make other inferences about the speaker
  - **Recognition:**
    - The system cannot determine anything more than the words spoken
      - I.e. cannot derive biometric or demographic info from the voice

- **Sharing:**
  - System can derive information for training from voice
  - The learned models do not reveal anything about whose voices were included

Requirement: Usage

- **Usage/inference:** Somehow process the data such that:
    - **Biometrics:**
        - The system cannot link to other sources of voice
        - The biometric is revocable if lost
        - The system cannot reverse engineer biometric to know more about the speaker
        - The “system” cannot make any inferences besides what is permitted
            - E.g. recognize speech, but not learn about the speaker’s ID/gender/other info
            - E.g. biometrically verify the speaker, but be unable to make other inferences about the speaker
    - **Recognition:**
        - The system cannot determine anything more than the words spoken
            - I.e. cannot derive biometric or demographic info from the voice

- **Sharing:**
    - System can derive information for training from voice
    - The learned models do not reveal anything about whose voices were included

Differential Privacy

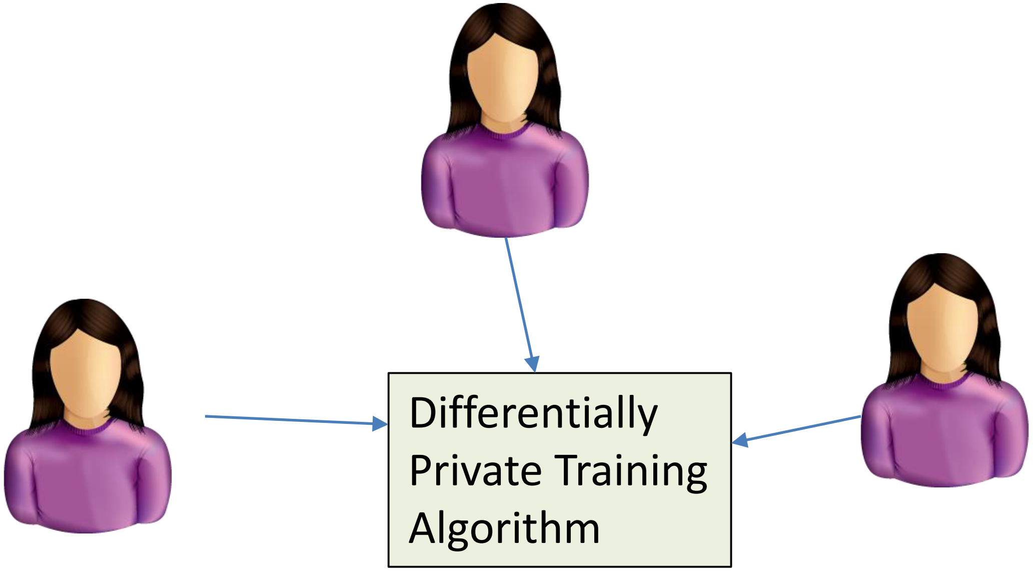
- A formalism that introduces noise in training so that inferences from aggregate data do not reveal individual data

- Differential privacy model:
  - *adjacent databases* D and D’ differing by one element
  - A randomized query M is differentially private if the probability that M produces response S from D is very close to the probability that it produces the same response S from D’

$$\left|\log \frac{P\,[M(D) \in S]}{P\,[M(D') \in S]}\right| \leqslant \epsilon \qquad \epsilon\text{-differential privacy}$$

<u>There is little or no increase in privacy risk if an individual chooses to contribute to the database</u>

Simple setting



Differentially Private Training Algorithm

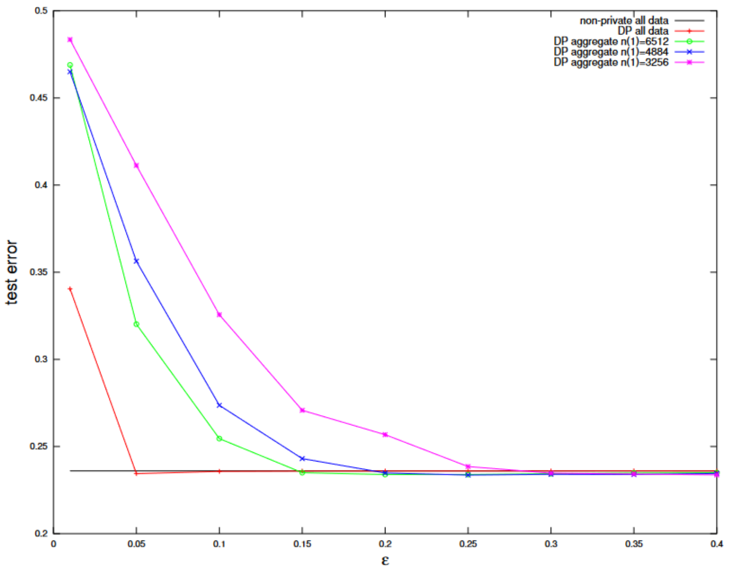
- Many parties contribute to the training of a pooled classifier
- DP used to ensure that every party is differentially private
  - Optimal strategy: Each participant locally computes classifier parameters and adds a tiny amount of noise before contributing it to the pool

Multi-party differential privacy: Excess Risk

$$\begin{aligned}\tilde{J}(\hat{\boldsymbol{w}}^s) \le \tilde{J}(\boldsymbol{w}^*) &+ \frac{2(K-1)^2(\lambda+1)}{2n_{(1)}^2\lambda^2} + \frac{4d^2(\lambda+1)}{n_{(1)}^2\epsilon^2\lambda^2}\log^2\left(\frac{d}{\delta}\right) \\ &+ \frac{4d(K-1)(\lambda+1)}{n_{(1)}^2\epsilon\lambda^2}\log\left(\frac{d}{\delta}\right) + \frac{16}{\lambda n}\left[32+\log\left(\frac{1}{\delta}\right)\right].\end{aligned}$$

- Everything to the right of the first “+” is excess risk
- Blows up inversely to $\frac{1}{\epsilon^2}$
  - For DP, ideally epsilon must be close to 0, but this will blow up the classifier’s performance

Differential privacy



Training a pooled softmax classifier

- Further left is more private
- 0.05 is not really private
    - ~5% probability
- Increased privacy destroys performance

Requirement: Usage

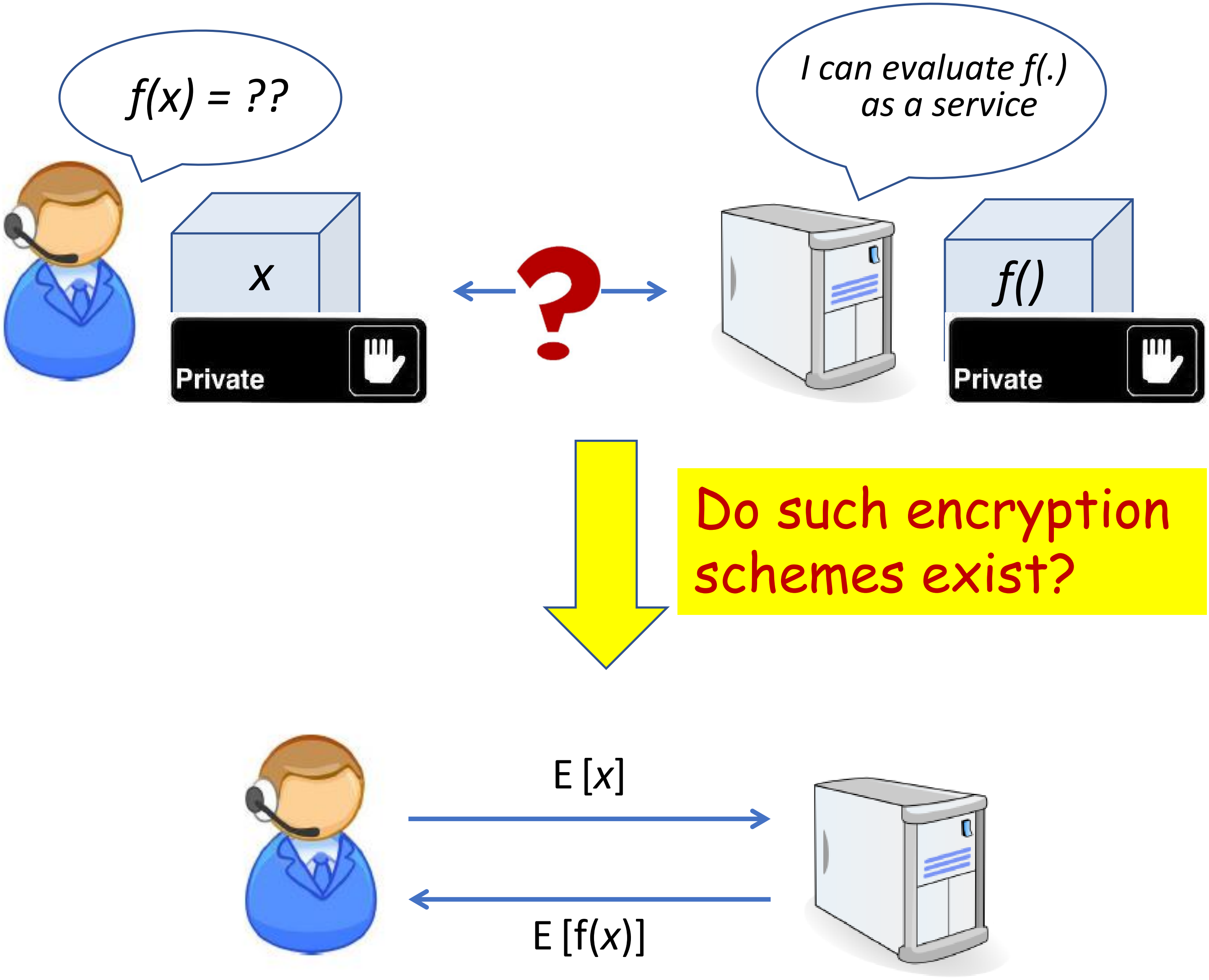
- **Usage/inference:** Somehow process the data such that:
  - **Biometrics:**
    - The system cannot link to other sources of voice
    - The biometric is revocable if lost
    - The system cannot reverse engineer biometric to know more about the speaker
    - The “system” cannot make any inferences besides what is permitted
      - E.g. recognize speech, but not learn about the speaker’s ID/gender/other info
      - E.g. biometrically verify the speaker, but be unable to make other inferences about the speaker
  - **Recognition:**
    - The system cannot determine anything more than the words spoken
      - I.e. cannot derive biometric or demographic info from the voice
- **Sharing:**
  - System can derive information for training from voice
  - The learned models do not reveal anything about whose voices were included

Requirement: Usage

- **Usage/inference:** Somehow process the data such that:
  - **Biometrics:**
    - The system cannot link to other sources of voice
    - The biometric is revocable if lost
    - The system cannot reverse engineer biometric to know more about the speaker
    - The “system” cannot make any inferences besides what is permitted
      - E.g. recognize speech, but not learn about the speaker’s ID/gender/other info
      - E.g. biometrically verify the speaker, but be unable to make other inferences about the speaker
  - **Recognition:**
    - The system cannot determine anything more than the words spoken
      - I.e. cannot derive biometric or demographic info from the voice

Can we simply encrypt the data and work off encrypted data?

Encrypting to hide information



Do such encryption schemes exist?

The approaches

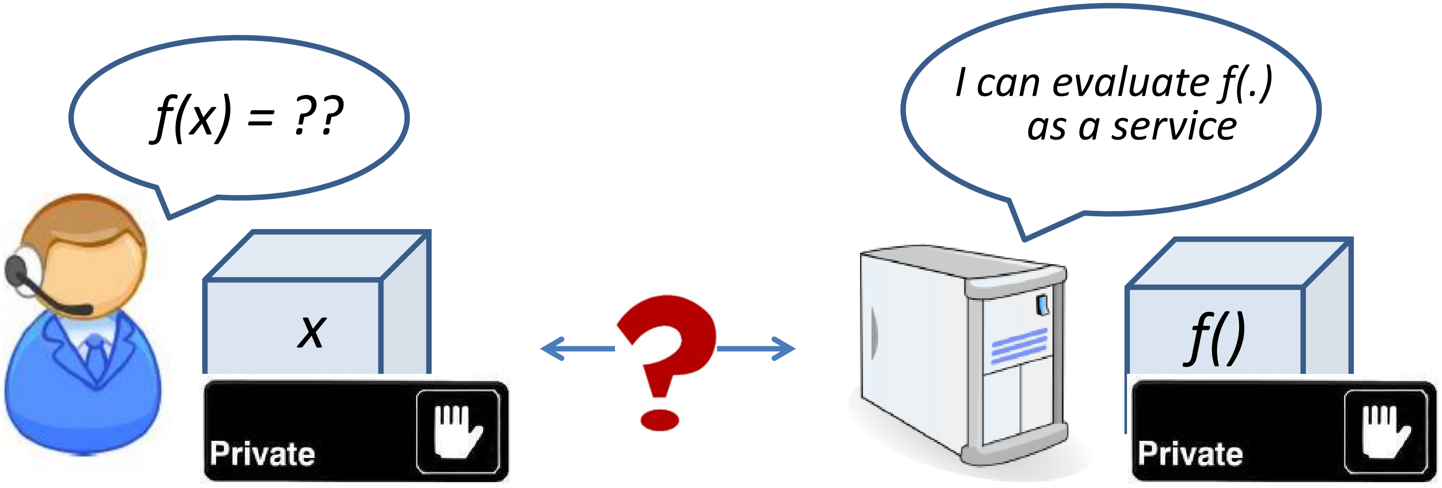
- **Homomorphic encryption**

Homomorphic Encryption

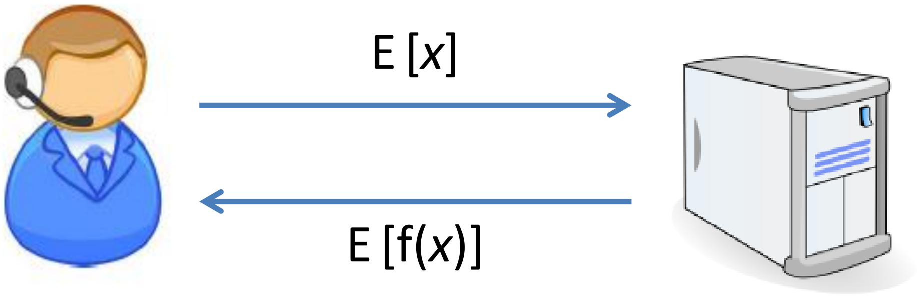
- Allows for operations to be performed on ciphertexts without requiring knowledge of corresponding plaintexts

$$E(x) \otimes E(y) = E(x \oplus y)$$

Homomorphic Encryption



$$E(x) \oplus E(y) = E(x \otimes y)$$



A “somewhat” homomorphic scheme

To encrypt a single bit b :

$$c = b + 2\overset{\text{n bit}}{p} + \overset{\text{N bit}}{kx}$$

- b is the message
- p is a random n-bit number
- x is any integer
- k is an N-bit key
- Note : “real” schemes work with integers over “ideal lattices”

A “somewhat” homomorphic scheme

$$c\ = b\ +\ 2p\ +\ kx$$

To decrypt :

$$b = (c\ mod\ k)mod\ 2$$

- $mod\ k$ “removes” kx
- $mod\ 2$ “removes $2p$”

A “somewhat” homomorphic scheme

$$c_1 = b_1 + 2p_1 + kx_1$$

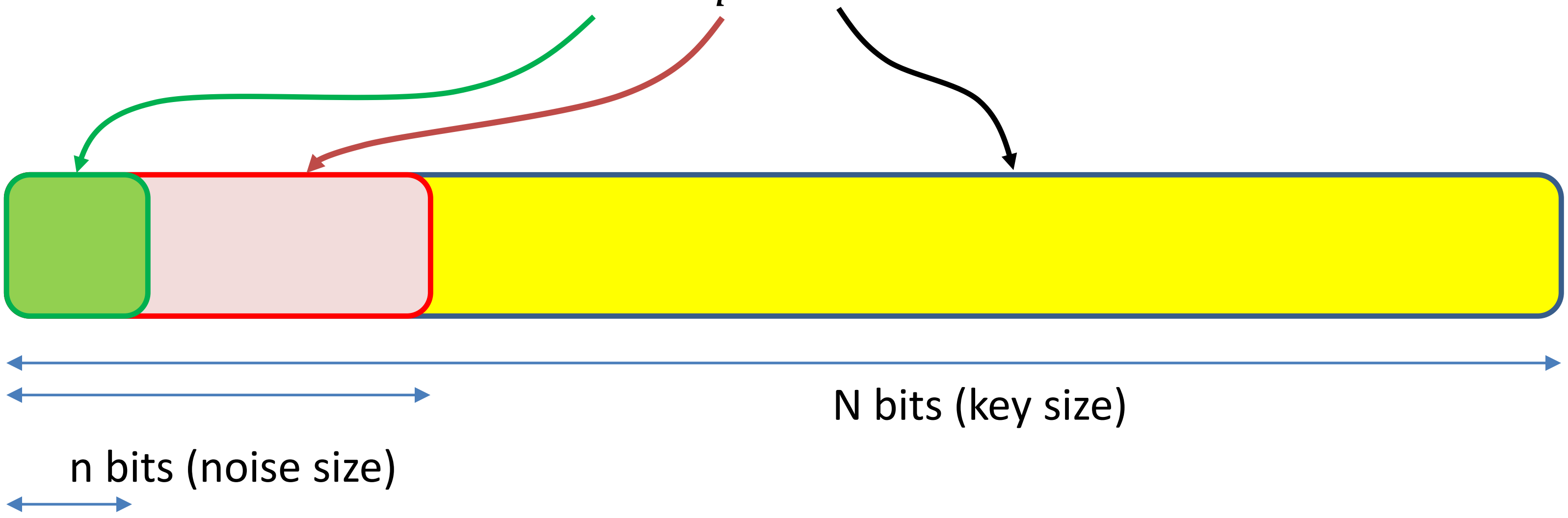
$$c_2 = b_2 + 2p_2 + kx_2$$

$$\begin{aligned} & c_1 + c_2 \\ & = b_1 + b_2 + 2(p_1 + p_2) + k(x_1 + x_2) \\ & = Enc(b_1 + b_2) \end{aligned}$$

$$\begin{aligned} c_1 c_2 & = b_1 b_2 + 2\mathcal{O}(p^2) + k\mathcal{O}(kx^2) \\ & = Enc(b_1 b_2) \end{aligned}$$

Somewhat homomorphic scheme

$$c = b + 2p + kx$$



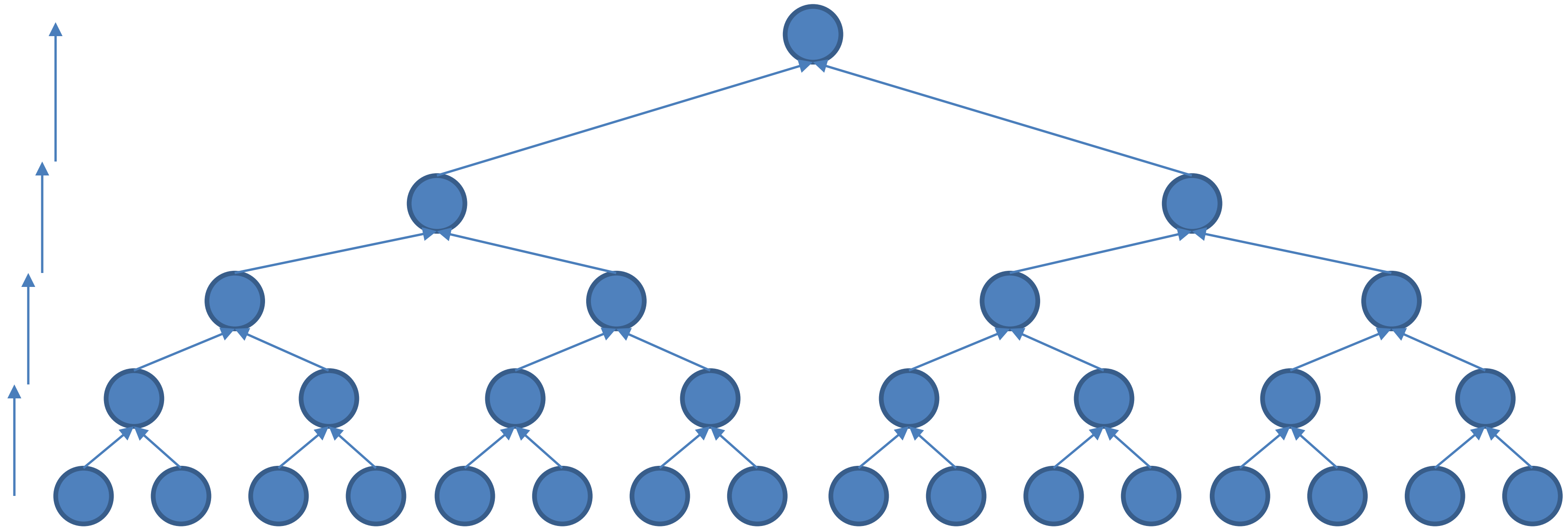
k bits message

- Decryption will fail if noise size > key size OR
- message size > noise size

Abstraction of typical computation

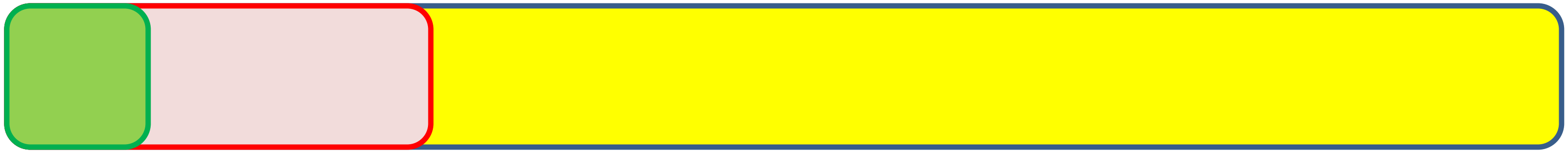
Each circle represents a variable
Each incoming pair of arrows represents a new variable obtained by multiplying two variables

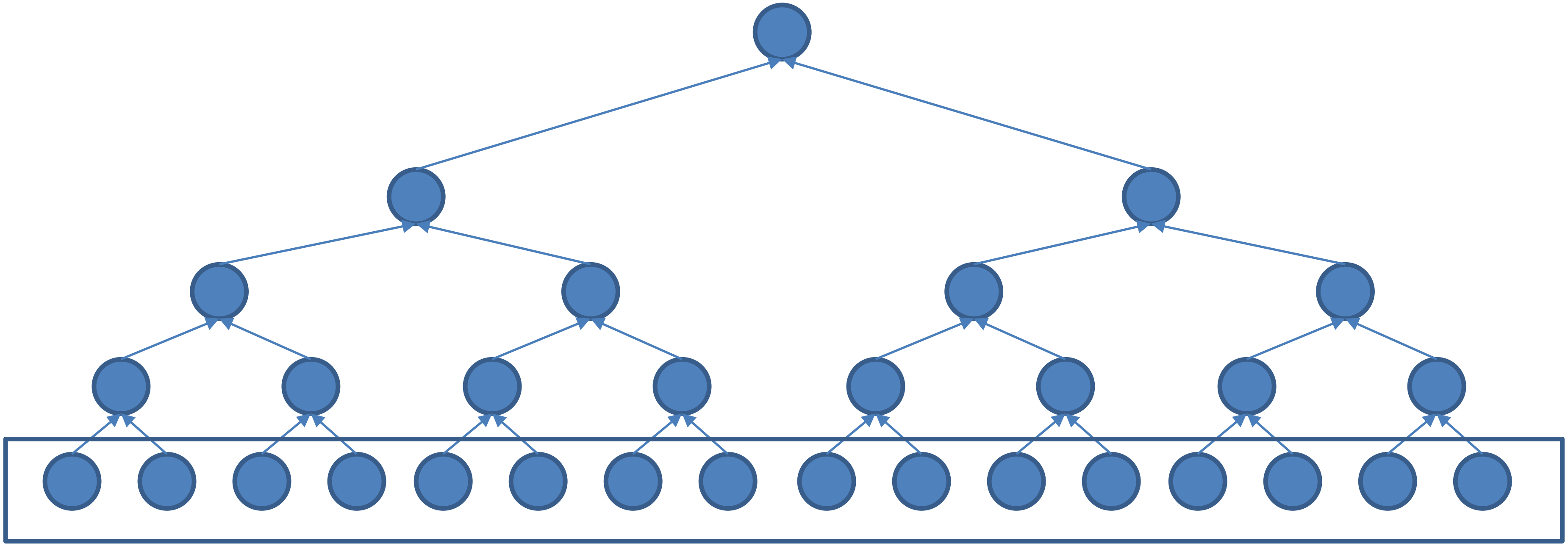
Actual graph will not be balanced
Can be hundreds of layers deep



Abstraction of typical computation

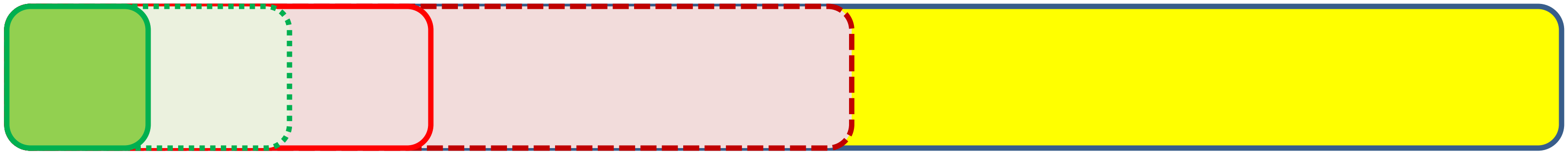
Each layer of multiplication (potentially) doubles the noise bits and the message bits
Pretty soon we will run out of key bits because the noise will exceed key size



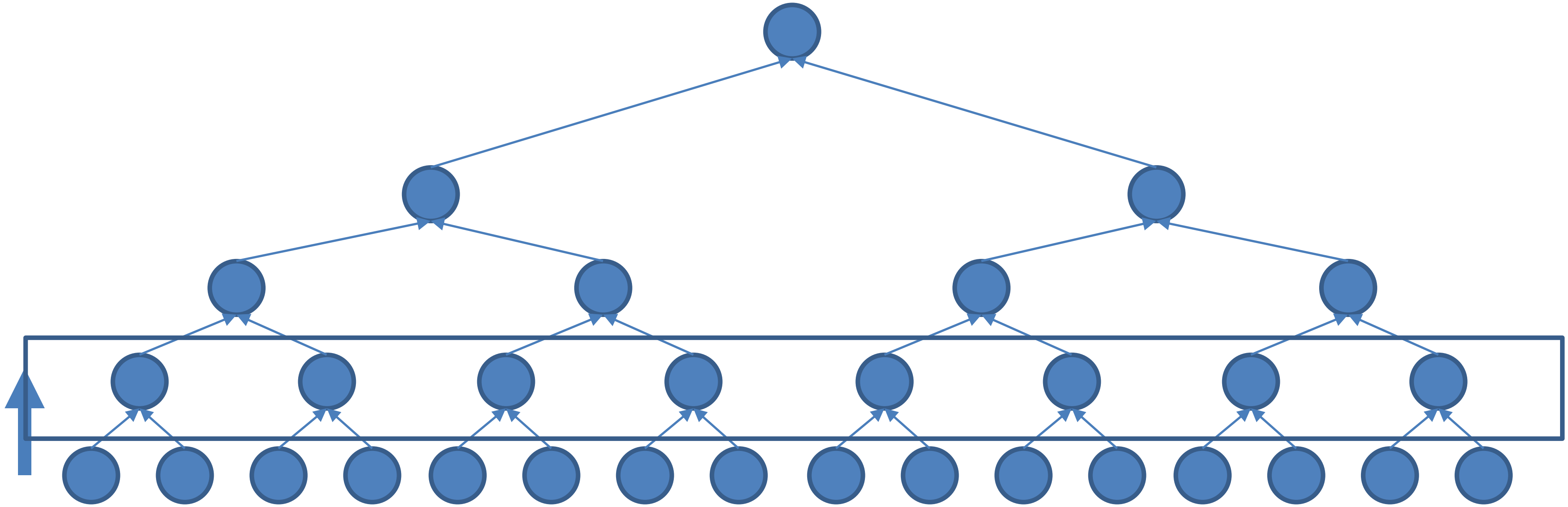


Abstraction of typical computation

Each layer of multiplication (potentially) doubles the noise bits and the message bits
Pretty soon we will run out of key bits because the noise will exceed key size

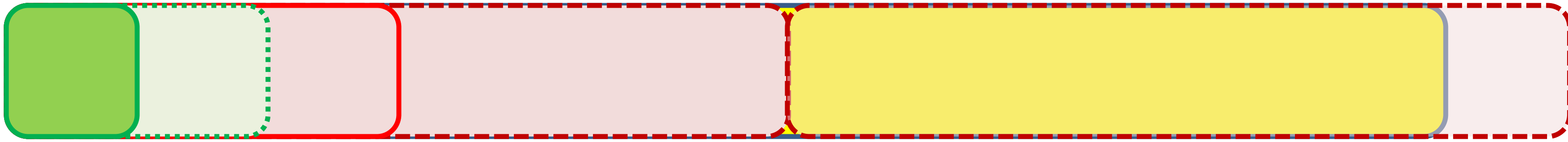


EACH LAYER OF COMPUTE DOUBLES NOISE SIZE

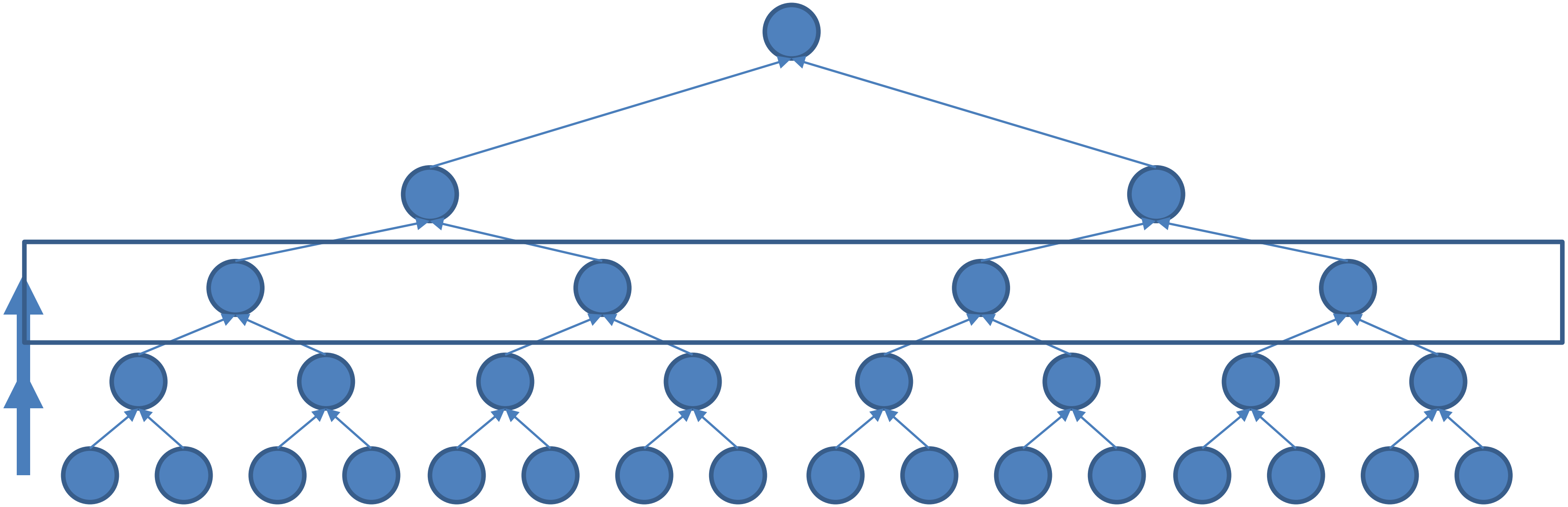


Abstraction of typical computation

Each layer of multiplication (potentially) doubles the noise bits and the message bits
Pretty soon we will run out of key bits because the noise will exceed key size

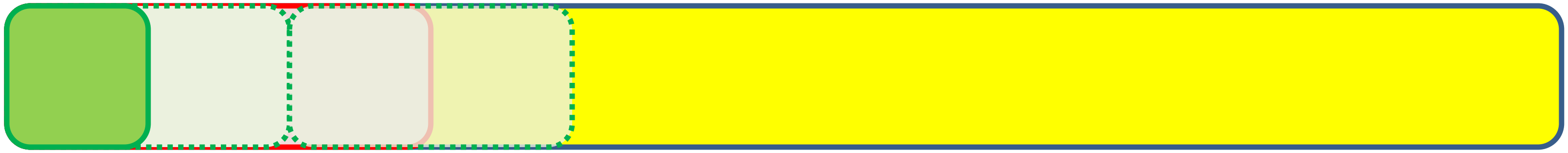


EACH LAYER OF COMPUTE DOUBLES NOISE SIZE

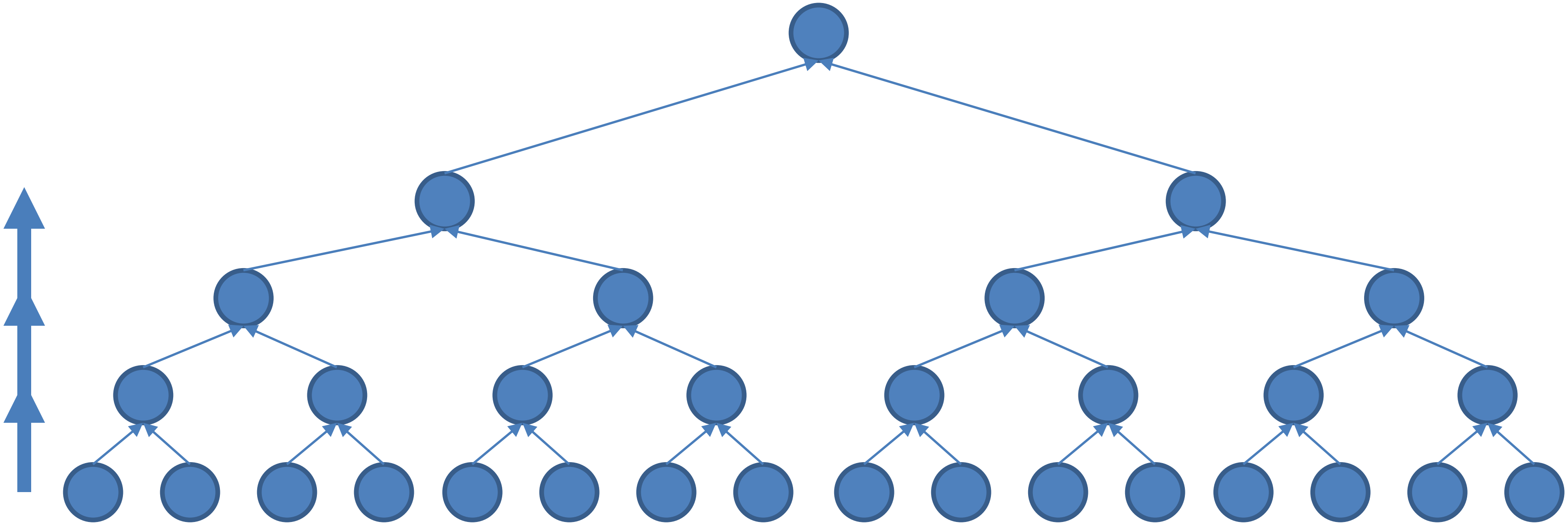


Abstraction of typical computation

Even if we manage not to increase the noise bits (as in newer schemes), we will run out of message bits



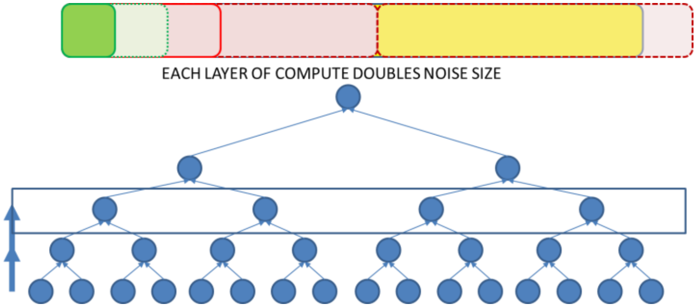
EACH LAYER OF COMPUTE DOUBLES MESSAGE SIZE



Homomorphic Encryption: Second issue

$$c_1 + c_2 = Enc(b_1 + b_2)$$

$$c_1 c_2 = Enc(b_1 b_2)$$

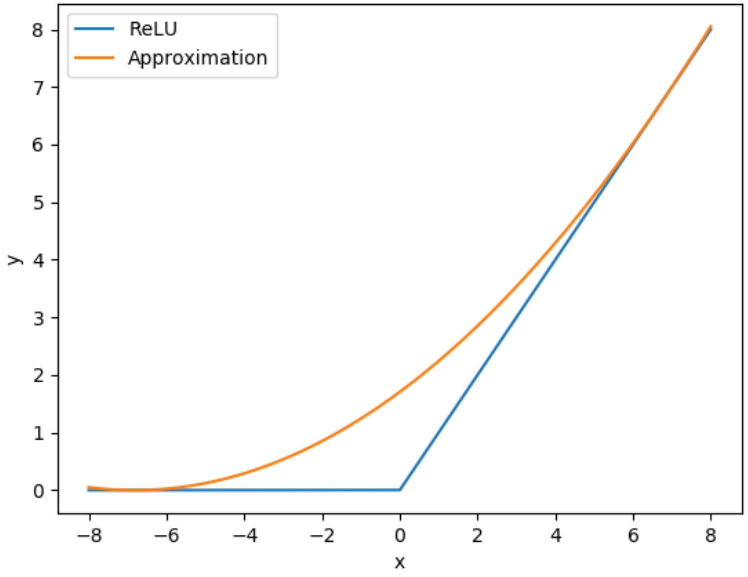


- Only allows additions and multiplications in encrypted domain
- Only permits computation of simple polynomials
  - Simple smooth curves
  - Need to approximate functions with corners or more complex functions with high-order polynomials
  - But those will quickly run out of bits

HE approximations to common activation functions

- Activation functions in NNs need to be approximated by polynomials.

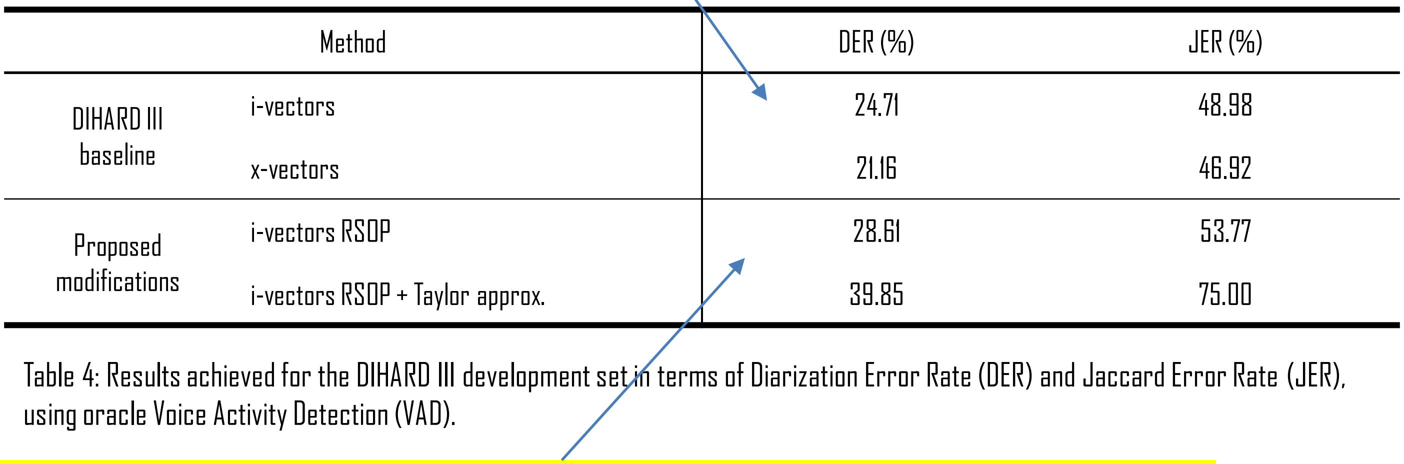
- Usually done through Chebyshev Polynomial approximations – provide good approximations on pre-determined intervals.



$$y = 0.037x^2 + 0.5x + 1.71$$

Results for a speech diarization task

Performance of speech diarization using “clear text” computation



Method		DER (%)	JER (%)
DIHARD III baseline	i-vectors	24.71	48.98
	x-vectors	21.16	46.92
Proposed modifications	i-vectors RSOP	28.61	53.77
	i-vectors RSOP + Taylor approx.	39.85	75.00

Table 4: Results achieved for the DIHARD III development set in terms of Diarization Error Rate (DER) and Jaccard Error Rate (JER), using oracle Voice Activity Detection (VAD).

Performance of speech diarization using HE with approximations

- **HE implementation takes ~30min to process 0.01 sec of speech.**

The approaches

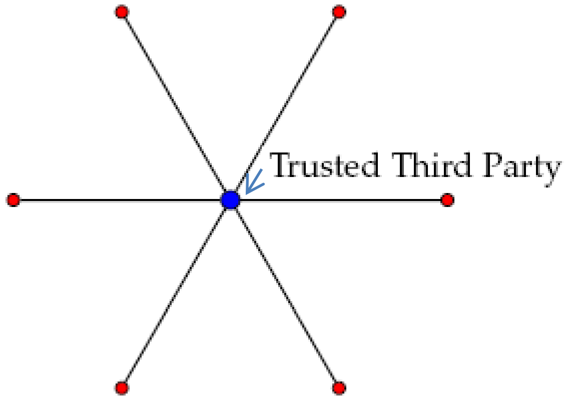
- ~~Homomorphic encryption~~
- **Secure multi-party computation**

Secure multiparty computation

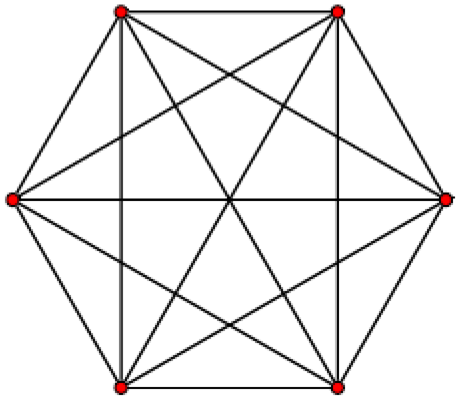
- Having one entity (the server) perform all the computation is not effective

- Can we do better if the parties *share* the computation?
  - Client and server each performs some part of the computation?
  - Only pre-specified entity retains the key that reveals the answer

SMC: Ideal Model and Real Model



Ideal Model



Real Model

Privacy-constraints trivially satisfied

Privacy-constraints satisfied if information learned by parties is indistinguishable from the ideal model case

SMC: Protocol Assumptions

- All parties know the protocol.
- There exists a public-key infrastructure and parties have access to public keys belonging to other parties
- Communication channels are reliable
- Parties (and potential adversaries) are computationally bounded in probabilistic polynomial time

Oblivious Transfer

- Bob has messages m_1 and m_K
- Alice wants m_n
  - But doesn't want Bob to know which message she wants

- Bob wants to let Alice have her desired message
  - But not the other ones

Oblivious Transfer

- Alice encodes

$$E = Paillier(n, k) \approx k^n$$

- Sends E and k to Bob

- For each message Bob computes $E_i = m_i\left(E\ k^{-i}\right)^r = m_i k^{r(n-i)}$
  - r is a random number
  - This will be m_i only for $i = n$, for the rest the result will be random

- Returns all E_i to Alice

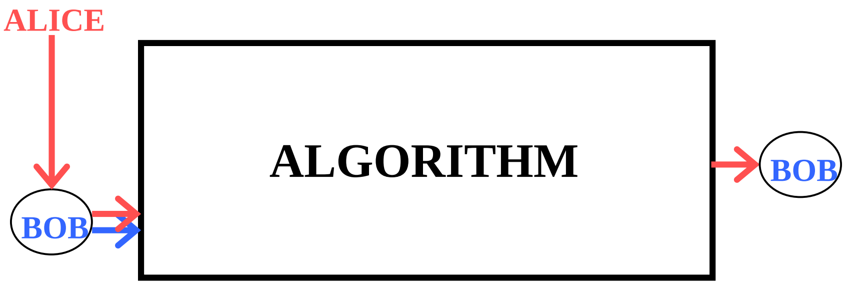
- Alice only reads E_n (the rest will be meaningless)

- Alice got her answer
  - Plus a lot of other gunk from the other unwanted data
- Bob never got to know what she wants

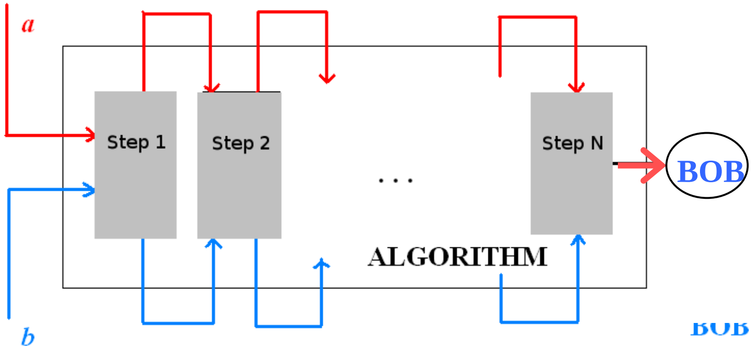
SMC

- SMC can be extended to compute *arbitrary* functions
  - Vector addition, Vector Innerproduct, Vector max, etc

SMC



- Conventional computation: User Alice sends data to system Bob
- Bob computes an algorithm



- SMC: Computation recast as a sequence of primitives
- Alice and Bob compute primitives via SMC
- Bob gets the result

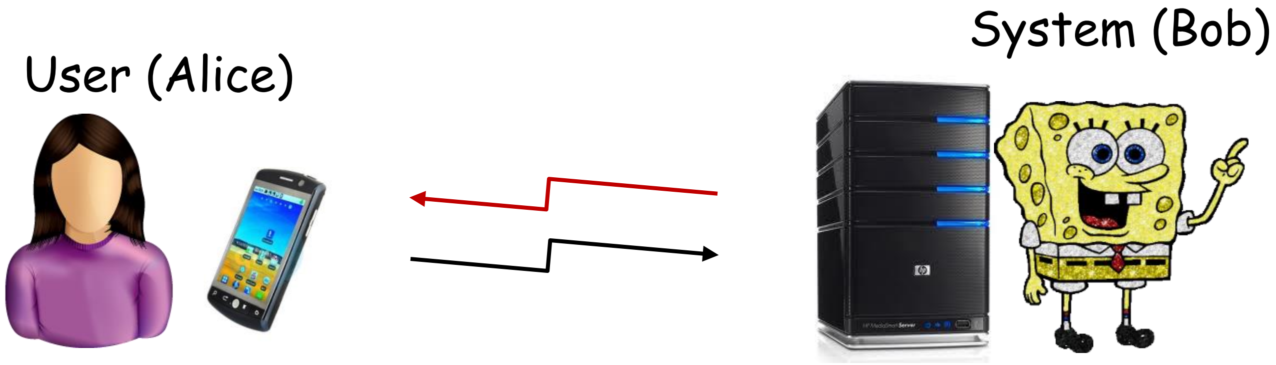
SMC

- SMC can be extended to compute *arbitrary* functions
  - Vector addition, Vector Innerproduct, Vector max, etc

- Problem:
  - Huge communication overhead

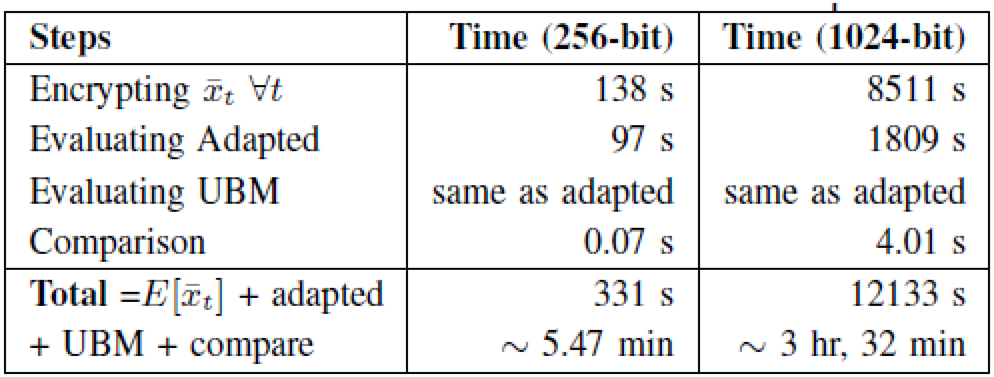
- Not really secure
  - Alice and Bob must both trust each other to do the right thing
    - "Honest but curious"
  - Adding additional layers of verification requires "zero-knowledge proofs" that can increase overhead by several orders of magnitude

SMC and Biometric Authentication



- User and system train template/model using SMC
  - System only learns encrypted models
- User and system use SMC to authenticate user
  - System never sees user’s data
  - System correctly authenticates user
- Performance: Identical to conventional non-private system

A Speaker Verification Task



Steps	Time (256-bit)	Time (1024-bit)
Encrypting $\bar{x}_t\ \forall t$	138 s	8511 s
Evaluating Adapted	97 s	1809 s
Evaluating UBM	same as adapted	same as adapted
Comparison	0.07 s	4.01 s
Total $=E[\bar{x}_t]$ + adapted + UBM + compare	331 s $\sim$ 5.47 min	12133 s $\sim$ 3 hr, 32 min

- Time taken to process 1 sec. of data
- Computation details:
  - Core-2 duo, 2 GHz
  - BGN cyrptosystem
    - Paillier an order of magnitude faster
  - Does not include communication overhead
- “Insecure” computation: insignificant
- Classification accuracies identical in secure and insecure versions

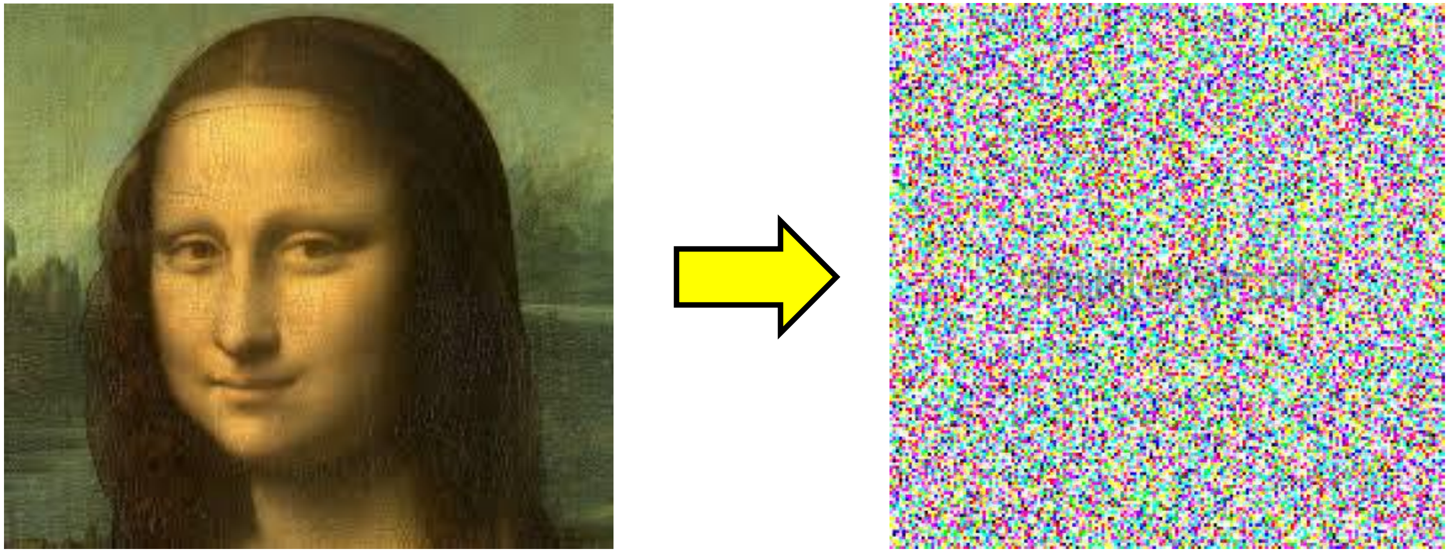
Factor of 100000 slow down for a relatively simple computation

More complex operations may be infeasible
Esp if we add zkps

The approaches

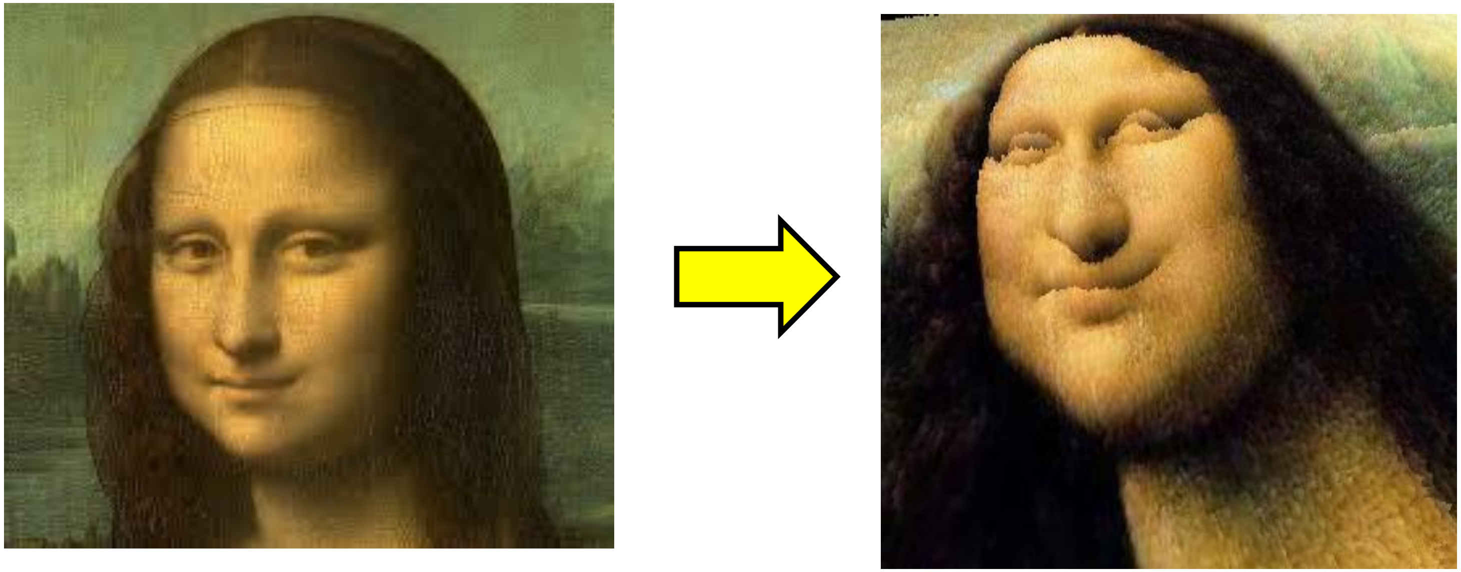
- ~~Homomorphic encryption~~
- ~~Secure multi-party computation~~
- **Transform-based solutions**

The Problem with Cryptographic Solutions



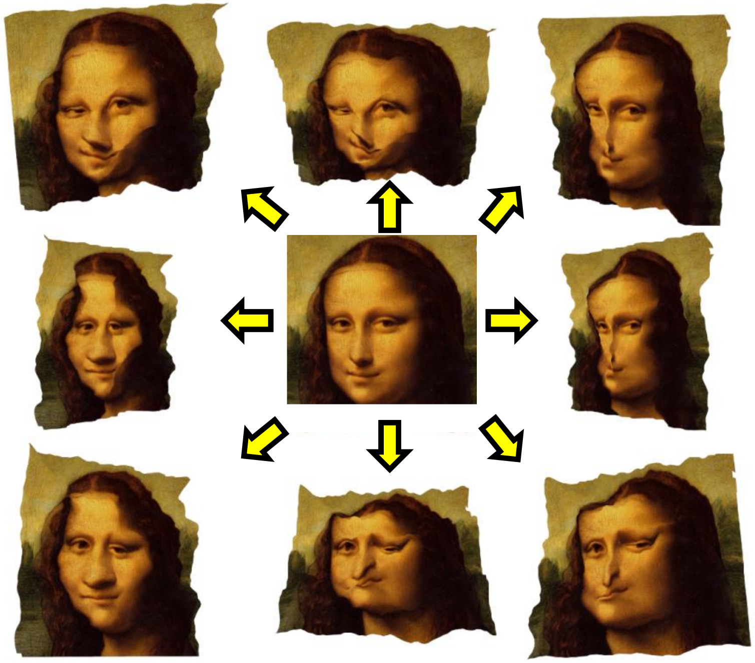
- Cryptographic techniques “hide” information by “shattering” the space
  - All notion of “neighborhood” is destroyed
  - Distance/neighborhood-based matching is impossible

Alternate solution: Transforms



- Instead of *shattering* the data space, distort it more continuously
  - *Transforms* that retain some measure of topological continuity
  - Permits continued estimation of neighborhood in transform domain

Requirement for Transforms

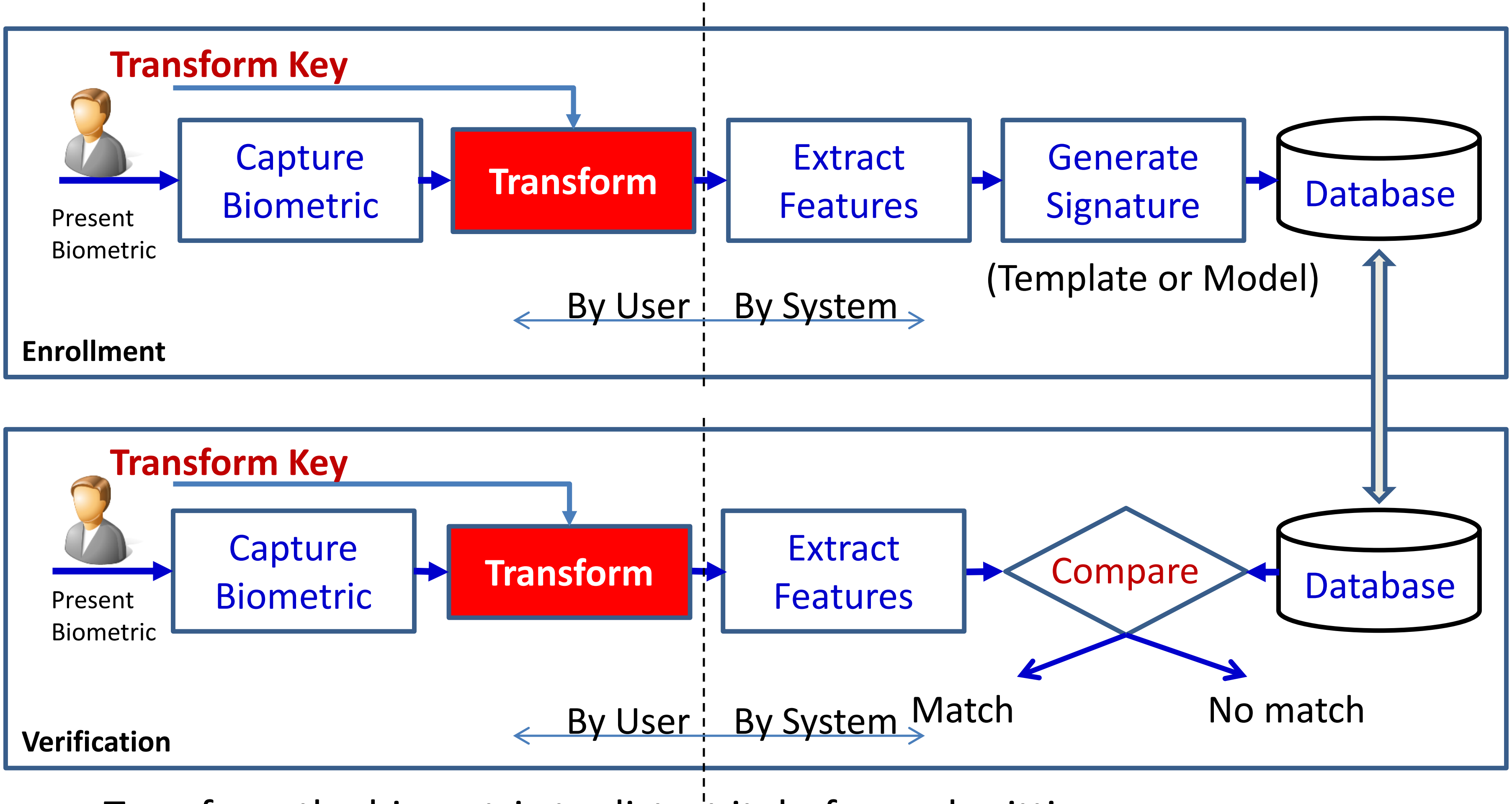


(c) entirelysubjective.com

- Transform must have user-specified parameter/key
  - Which controls the actual nature of the transform

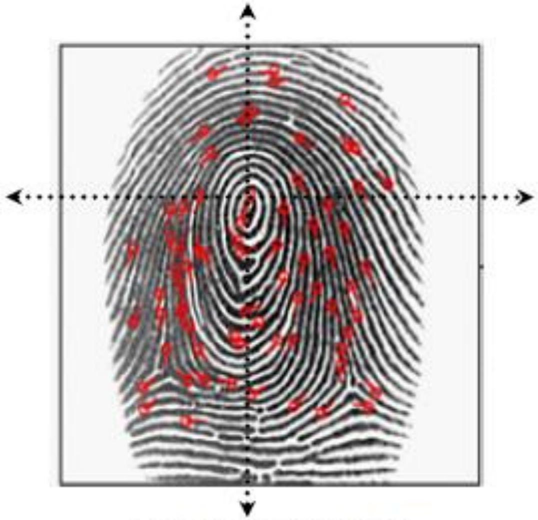
$$\boldsymbol{Y} = \boldsymbol{T}(\boldsymbol{X}, K)$$

Transform-based approach

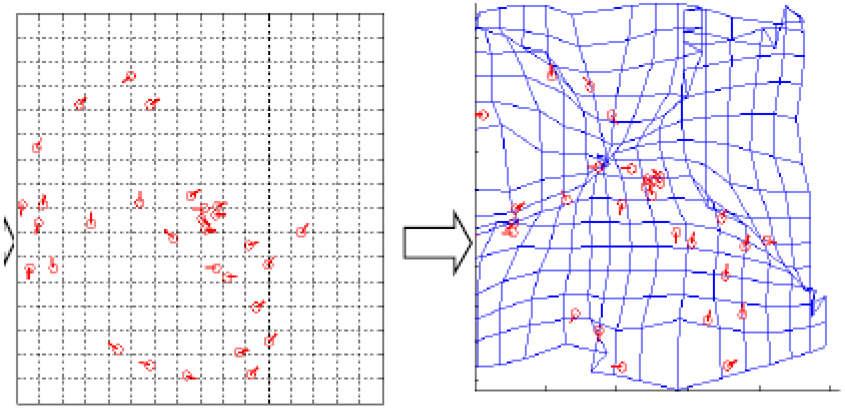


- Transform the biometric to distort it, before submitting
  - With a transform that has a user-specified transform parameter
- **Perform comparison over transformed biometric**

Transforms example: Distorting Fingerprints



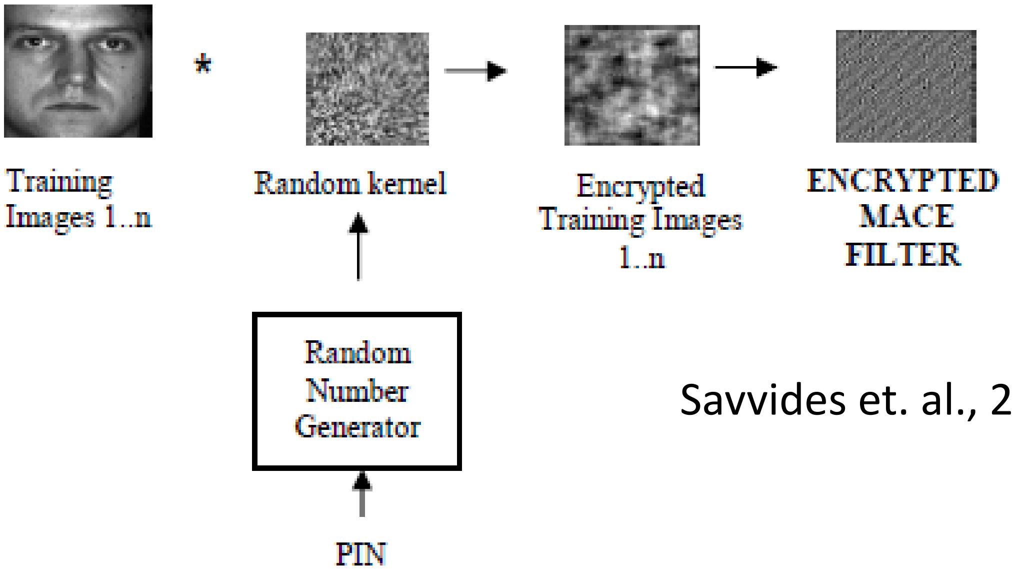
Feature Extraction



Surface Folding Transformation

- Ratha et. al., 2006
- Distort the surface of the fingerprint through a “crumpling” transform
- Example shown: Functional transform (transform obtained through a mathematical function)
  - The actual shape of the distortion depends on the user-specified key

Face (with random filters)



Savvides et. al., 2004

- Generate a random filter using a user-specified seed
  - Seed is user’s transform key
- Convolve image using random filter

Random Projections

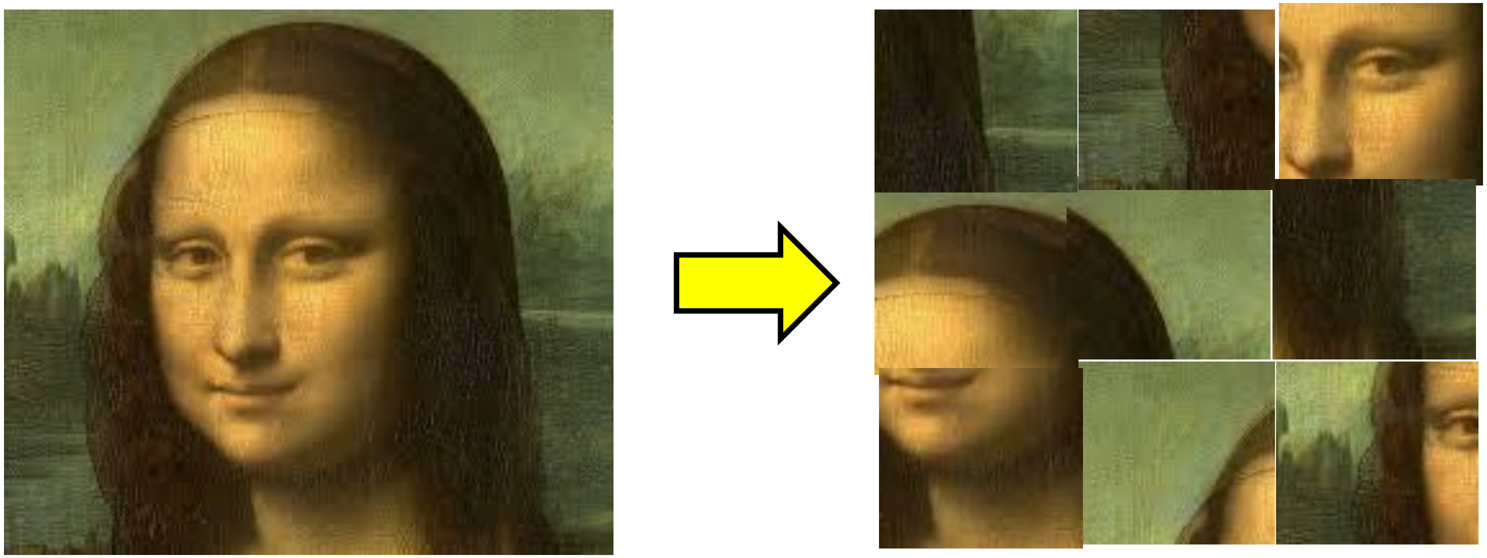
- Several authors have proposed the use of random projections

$$\boldsymbol{Y} = \boldsymbol{g}(\boldsymbol{\Phi X})$$

  - $\boldsymbol{X}$ is true biometric
  - $\boldsymbol{\Phi}$ is a random projection matrix (Key)
  - $\boldsymbol{Y}$ is transformed biometric
  - $\boldsymbol{g}$ is typically a quantization of some kind

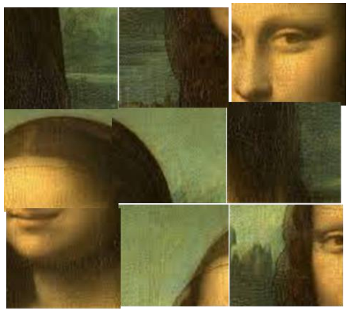
- Applicable to many different biometrics
- Can greatly increase false rejection or false alarm

Shuffling



- Shuffle the data

Problem



	x_1	x_2	x_3	x_4	x_5	x_6
x_1						
x_2						
x_3						
x_4						
x_5						
x_6						

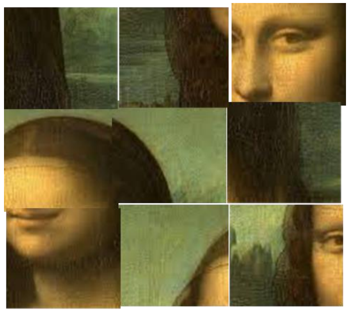
- Random transforms prevent the system from recovering the original biometric from a single instance
  - But do allow the distance between any pair of instances to be computed (used for comparison)

- Given a collection of instance, permit computation of a full matrix of pair-wise distances
- This matrix can be “inverted” to create an isometric map of the original space

The approaches

- ~~Homomorphic encryption~~
- ~~Secure multi-party computation~~
- ~~Transform-based solutions~~
- **Hashing-based solutions**

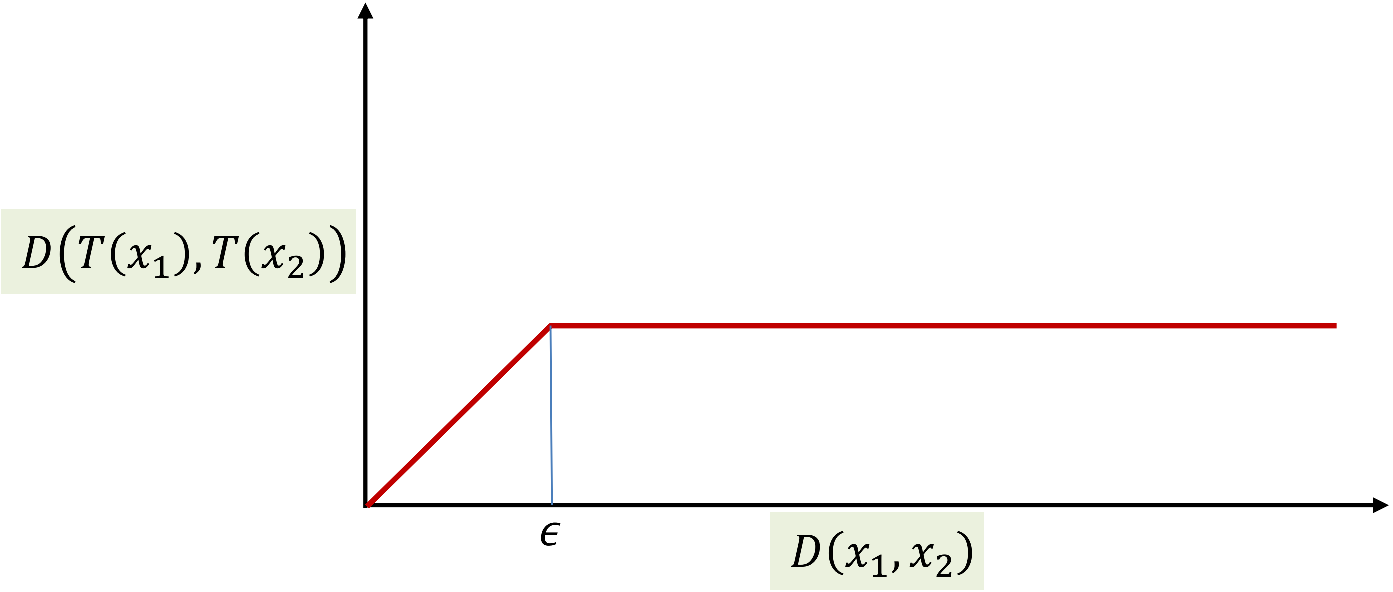
Problem



	x_1	x_2	x_3	x_4	x_5	x_6
x_1						
x_2						
x_3						
x_4						
x_5						
x_6						

- Random transforms prevent the system from recovering the original biometric from a single instance
  - But do allow the distance between **any** pair of instances to be computed (used for comparison)
- Given a collection of instance, permit computation of a full matrix of pair-wise distances
- This matrix can be "inverted" to create an isometric map of the original space

Restricting reveal of distance



- Is there a transform that will only reveal the distance in a limited way?
  - $D(T(x_1),T(x_2))$ reveals $D(x_1,x_2)$ if $D(x_1,x_2) < \epsilon$, but not otherwise

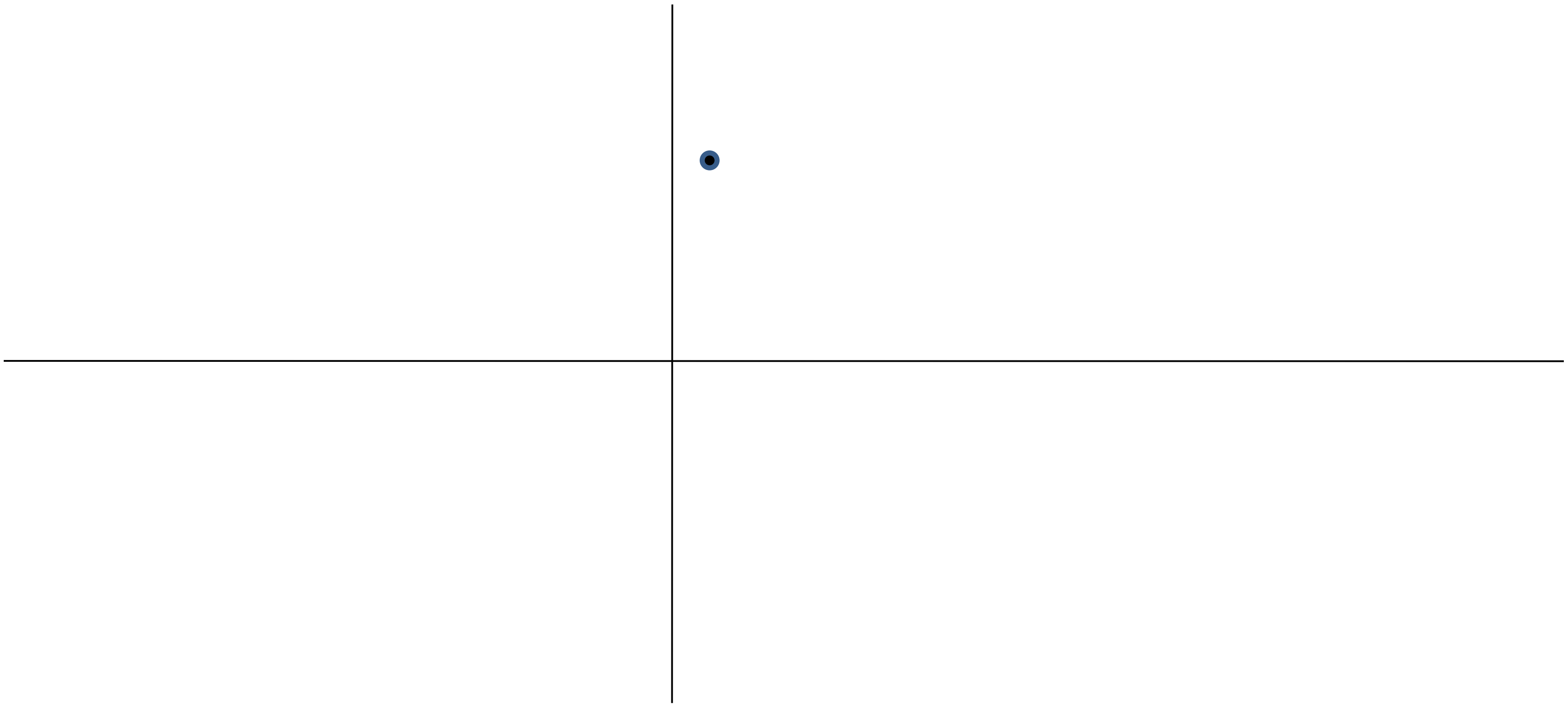
Limited-leakage Hashes

$$Q_k(\mathbf{x}) = \left\lfloor \varphi^\top \mathbf{x} + U \right\rfloor \pmod k$$

with $\varphi \sim \mathcal{N}\left(\mathbf{0}, \frac{1}{\delta^2} I_N\right)$ and $U \sim \mathrm{unif}(0, k)$ independent.

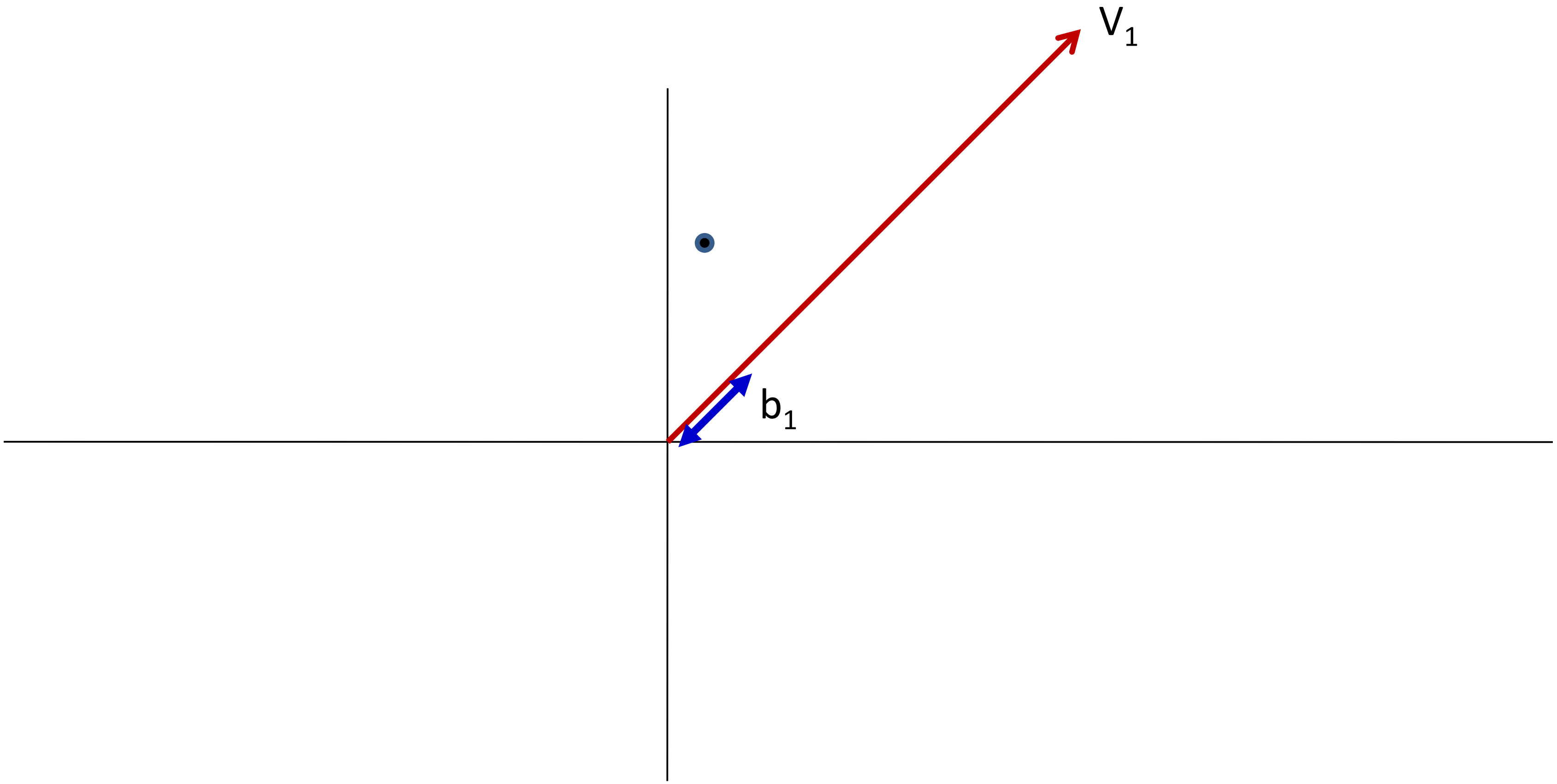
- Banded LSH

Euclidean LSH



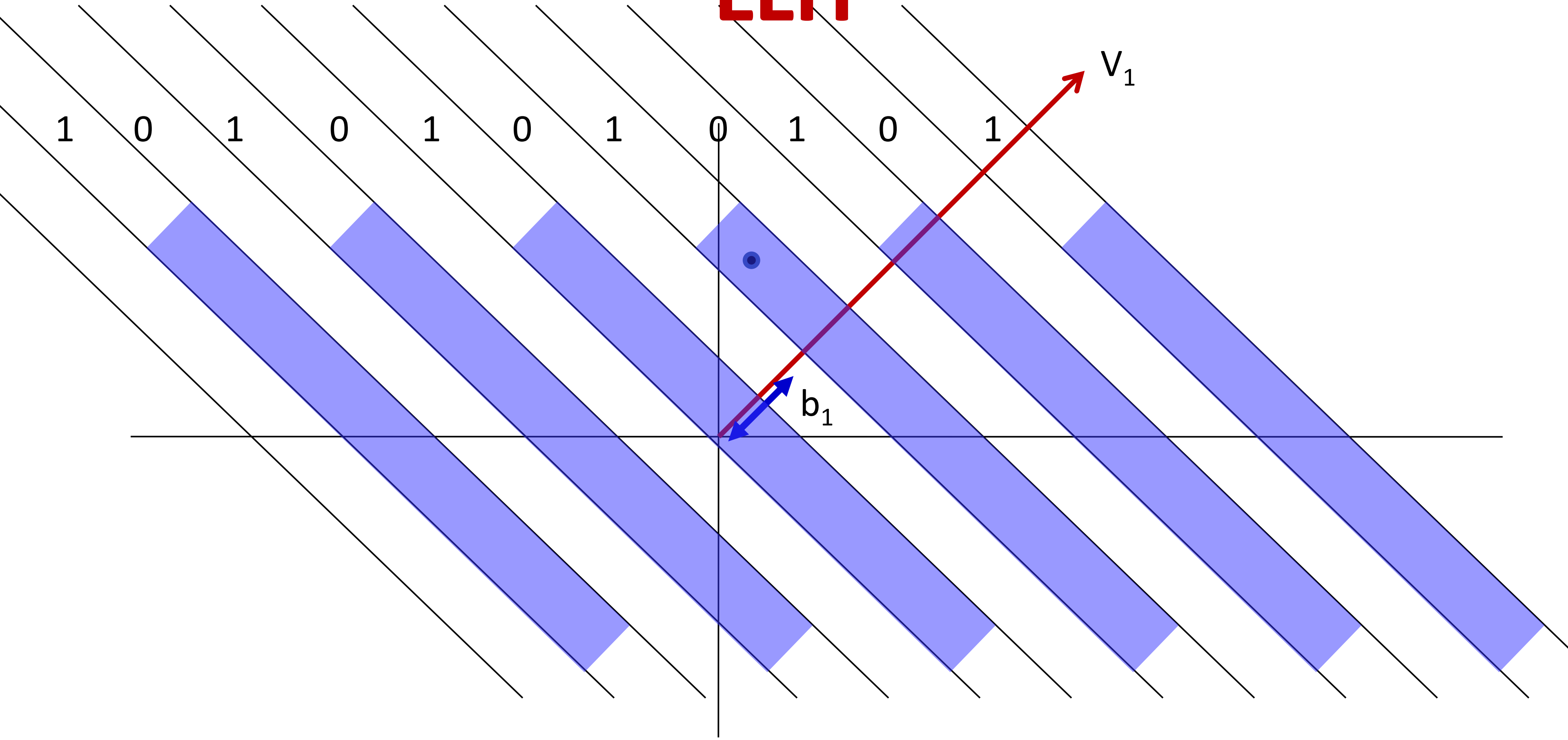
- A 2-D example

Euclidean LSH



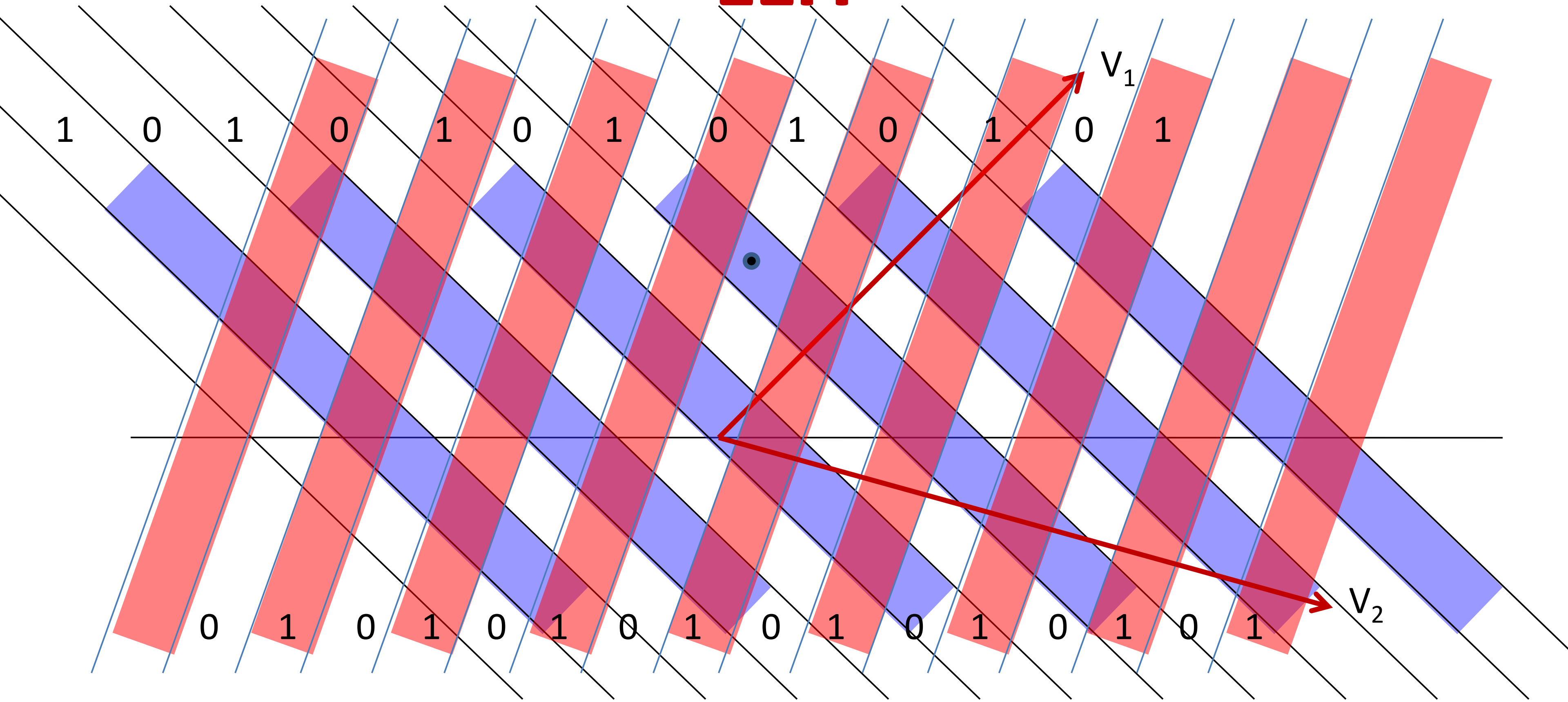
- A 2-D example
- The first component in the hash key: $h_1(X)$

LLH



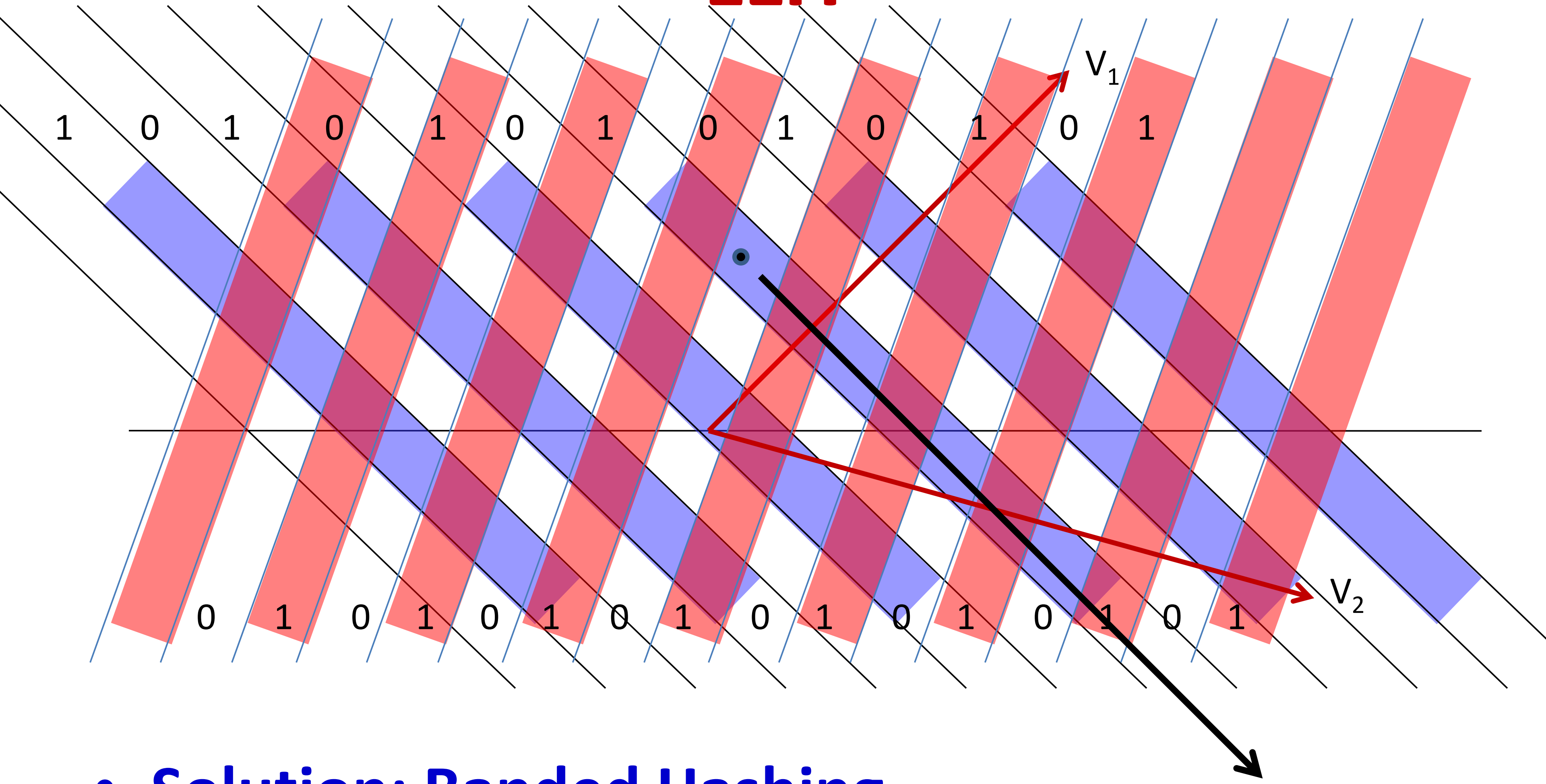
- A 2-D example
- The first component in the hash key : $h_1(X) = 1$

LLH



- **Solution: Banded Hashing**
  - Euclidean LSH with binary output Q(X)=[1,1]

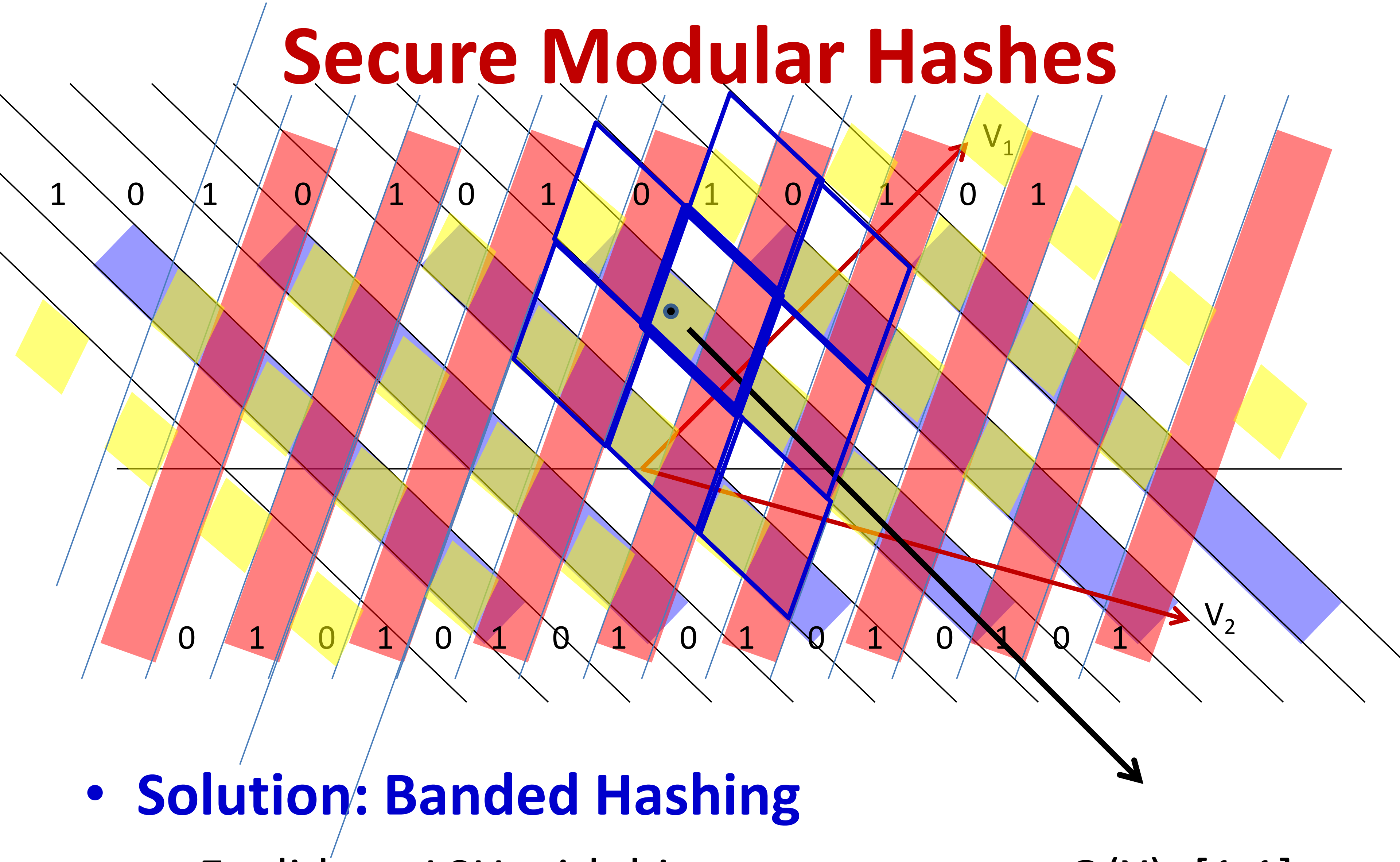
LLH



- **Solution: Banded Hashing**
  - Euclidean LSH with binary output

Q(X)=[1,1]

Secure Modular Hashes

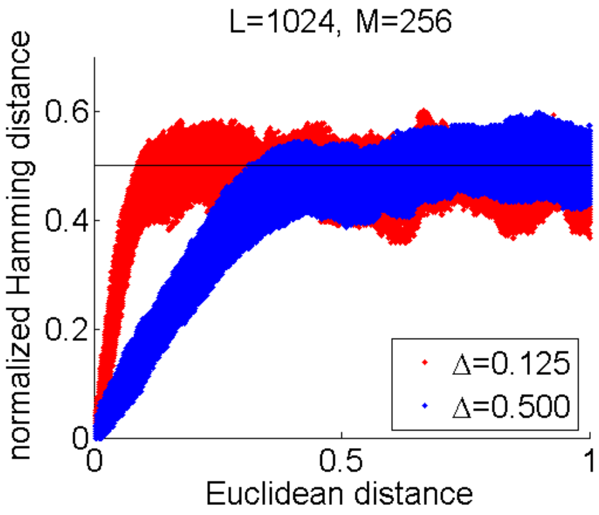


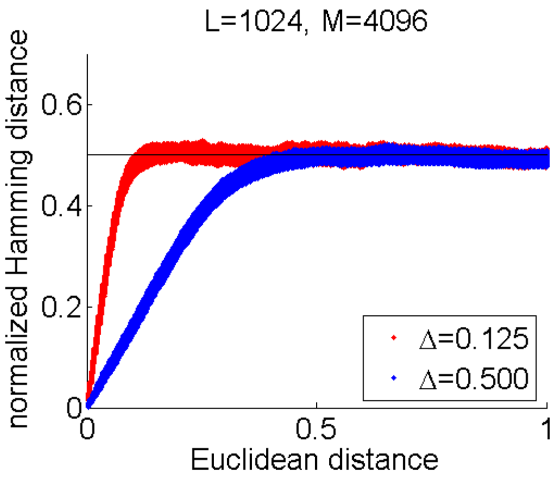
- **Solution: Banded Hashing**
  - Euclidean LSH with binary output

Q(X)=[1,1]

LLH

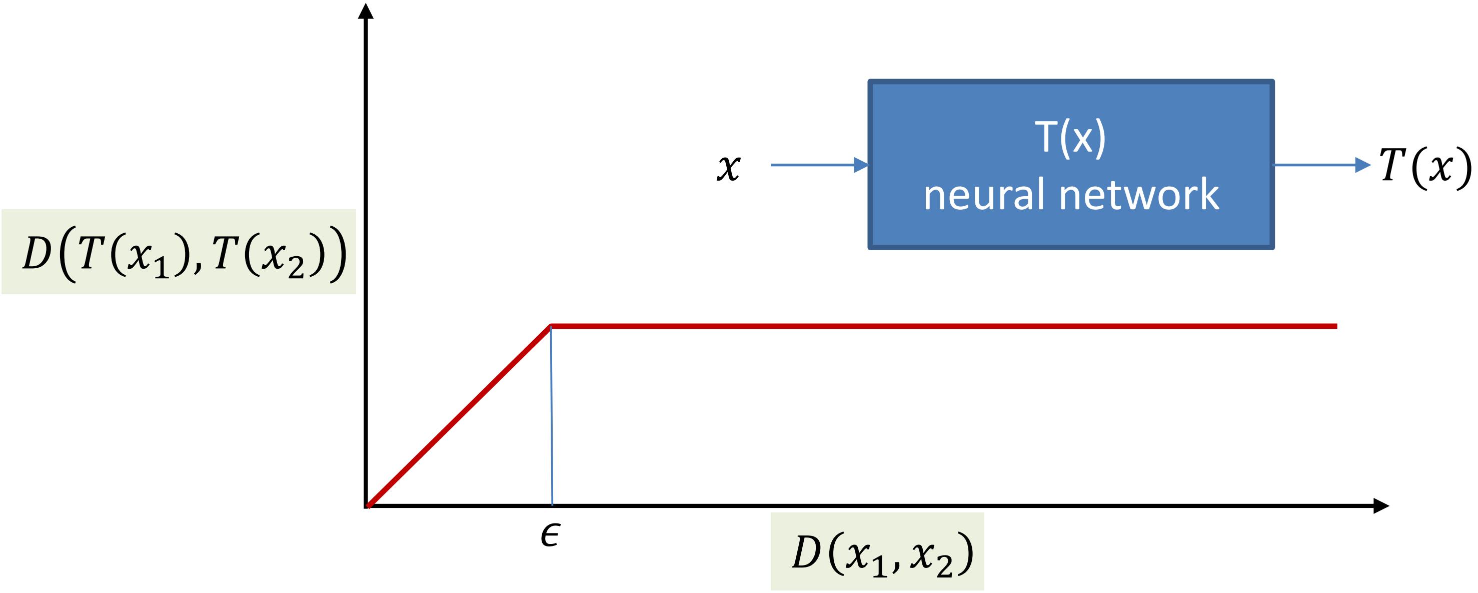
Simulations: L-dimensional vectors, M bit hashes





- Plot of Hamming(Q(X),Q(Y)) vs Euclidean d(X,Y) for different values of Δ, and different numbers of bits in Q(X)

Restricting reveal of distance



- $T(x)$ can even be learned
- For multi-attribute inputs, can be optimized to reveal distances only about specific attributes of the input, but not others
  - E.g for voice, can reveal limited information about id, but have lower sensitivity to age/emotion…

On a speaker verification task

- Conventional verification using insecure data
  - Equal Error Rate = 0.33%

- Classification with “private” SMH
  - Equal Error Rate = 0.5%

- Insignificant difference
- Extra computation: O(< 1ms)

Requirement: Usage

- **Usage/inference:** Somehow process the data such that:
  - **Biometrics:**
    - The system cannot link to other sources of voice
    - The biometric is revocable if lost
    - The system cannot reverse engineer biometric to know more about the speaker
    - The "system" cannot make any inferences besides what is permitted
      - E.g. recognize speech, but not learn about the speaker's ID/gender/other info
      - E.g. biometrically verify the speaker, but be unable to make other inferences about the speaker
  - **Recognition:**
    - The system cannot determine anything more than the words spoken
      - I.e. cannot derive biometric or demographic info from the voice

Only solved this

- **Sharing:**
  - System can derive information for training from voice
  - The learned models do not reveal anything about whose voices were included

Requirement: Usage

- **Usage/inference:** Somehow process the data such that:
  - **Biometrics:**
    - The system cannot link to other sources of voice
    - The biometric is revocable if lost
    - The system cannot reverse engineer biometric to know more about the speaker
    - The "system" cannot make any inferences besides what is permitted
      - E.g. recognize speech, but not learn about the speaker's ID/gender/other info
      - E.g. biometrically verify the speaker, but be unable to make other inferences about the speaker
  - **Recognition:**
    - The system cannot determine anything more than the words spoken
      - I.e. cannot derive biometric or demographic info from the voice
- **Sharing:**
  - System can derive information for training from voice
  - The learned models do not reveal anything about whose voices were included

No cryptographic solution feasible

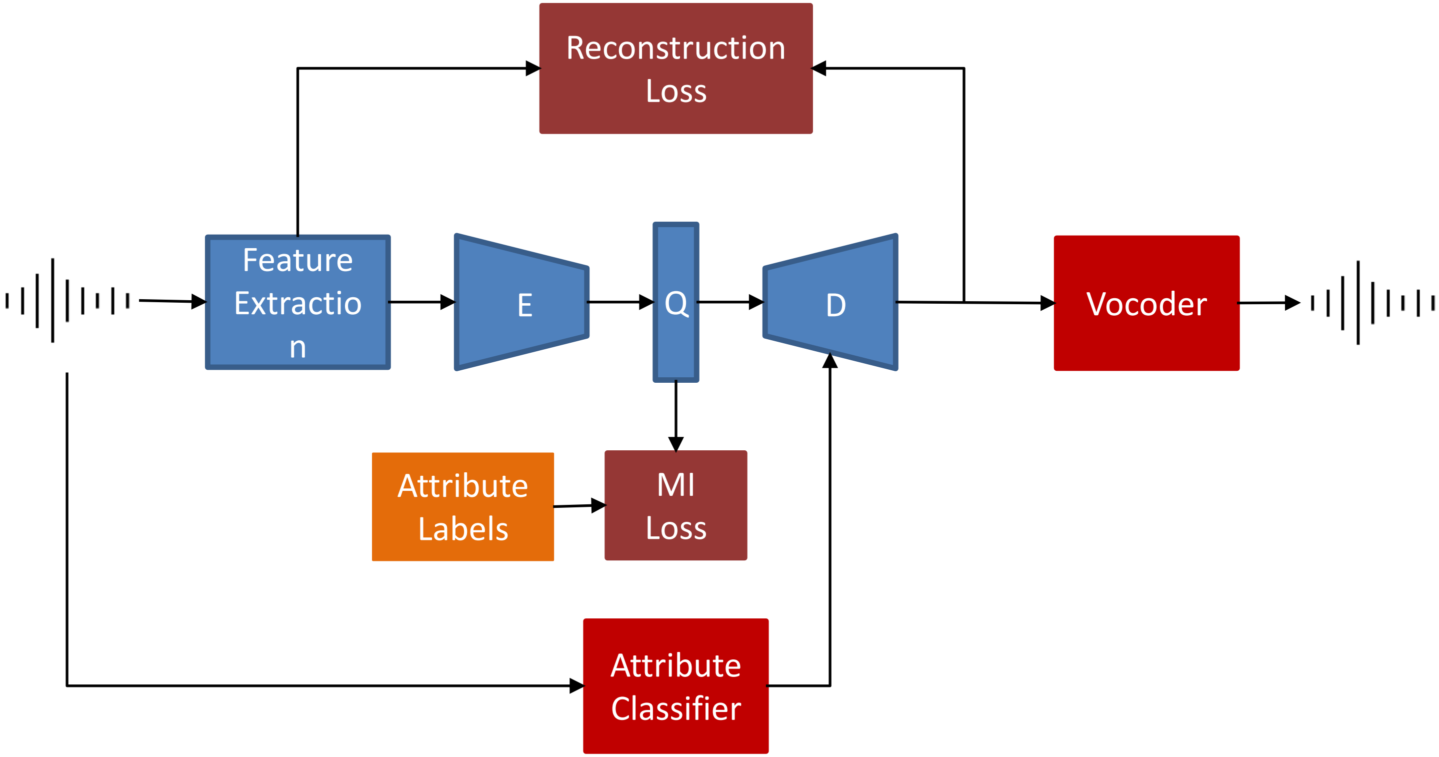
Requirement: Usage

- **Usage/inference:** Somehow process the data such that:
  - **Biometrics:**
    - The system cannot link to other sources of voice
    - The biometric is revocable if lost
    - The system cannot reverse engineer biometric to know more about the speaker
    - The "system" cannot make any inferences besides what is permitted
      - E.g. recognize speech, but not learn about the speaker's ID/gender/other info
      - E.g. biometrically verify the speaker, but be unable to make other inferences about the speaker
  - **Recognition:**
    - The system cannot determine anything more than the words spoken
      - I.e. cannot derive biometric or demographic info from the voice

Solution: Explicitly elide "sensitive" information from recordings

For this we must turn to neural networks with adversarial training

Attribute filtering via Mutual Information Minimization



Output of encoder must be incapable of classifying sensitive attribute

The filtering approach

- Can result in zero degradation of speaker verification performance
- Minimal degradation of speech recognition performance
- But are we *done?*
  - No
  - Can only eliminate one attribute with some success
  - Increasing the number of sensitive attributes to remove results in decreasing efficiency of elision for each atribute

Work in progress

- No specific solution is complete

- Efficient solutions are not effective and vice versa
- Is actually currently a very active area of research

- Our story so far is one of few successes
    - But lots of fun maths
        - Fwiw

The Abrupt Stop