Automated Side Channel Analysis of Media Software with Manifold Learning

Yuanyuan Yuan, Qi Pang, Shuai Wang

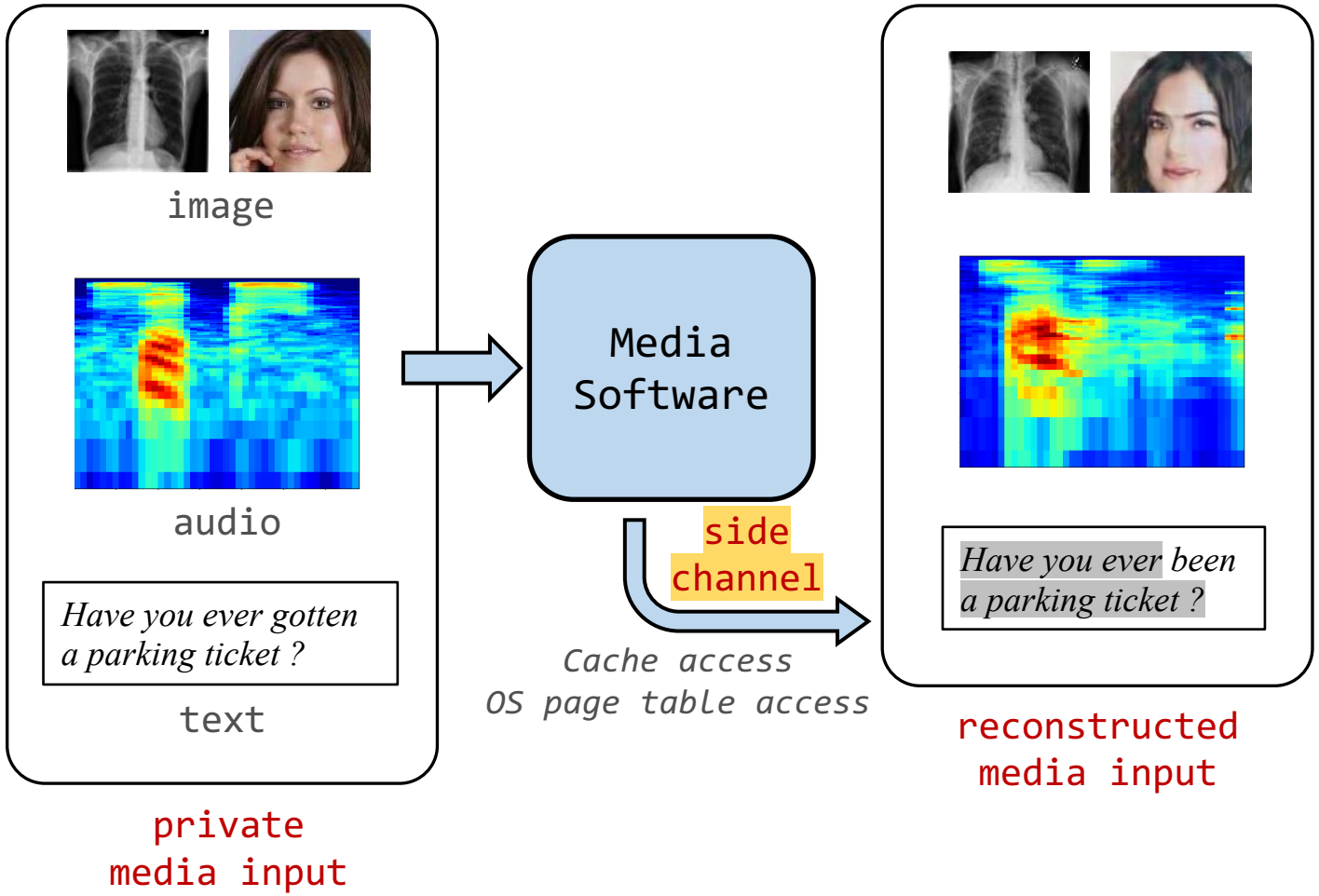
The Hong Kong University of Science and Technology

USENIX Security 2022

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

usenix
THE ADVANCED COMPUTING SYSTEMS ASSOCIATION

Threat Model & Overview



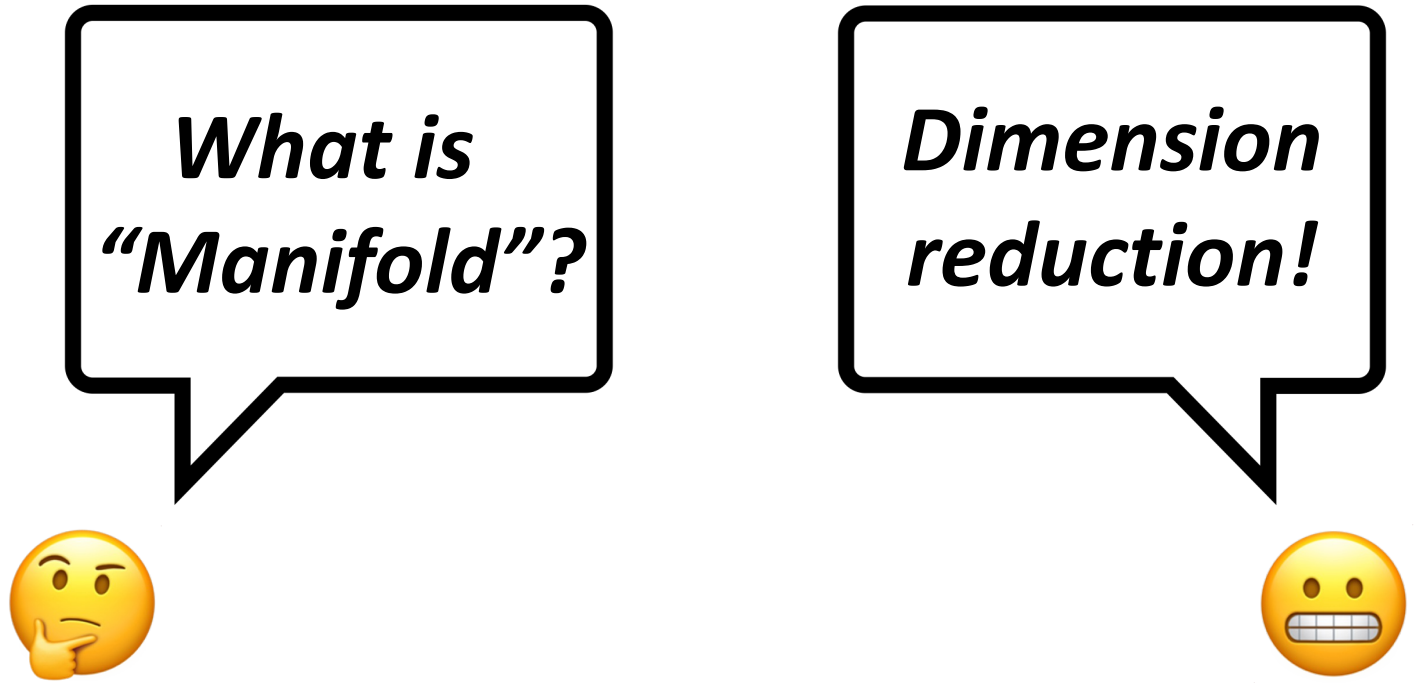
- Standard trace-based attack
- Executables
- Log side channels via
  1) Intel Pin (for "debugging")
  2) Prime & Probe

- ❑ **Reconstruct**
  private media inputs
- ❑ **Localize**
  vulnerable program points
- ❑ **Mitigate**
  with *perception blinding*

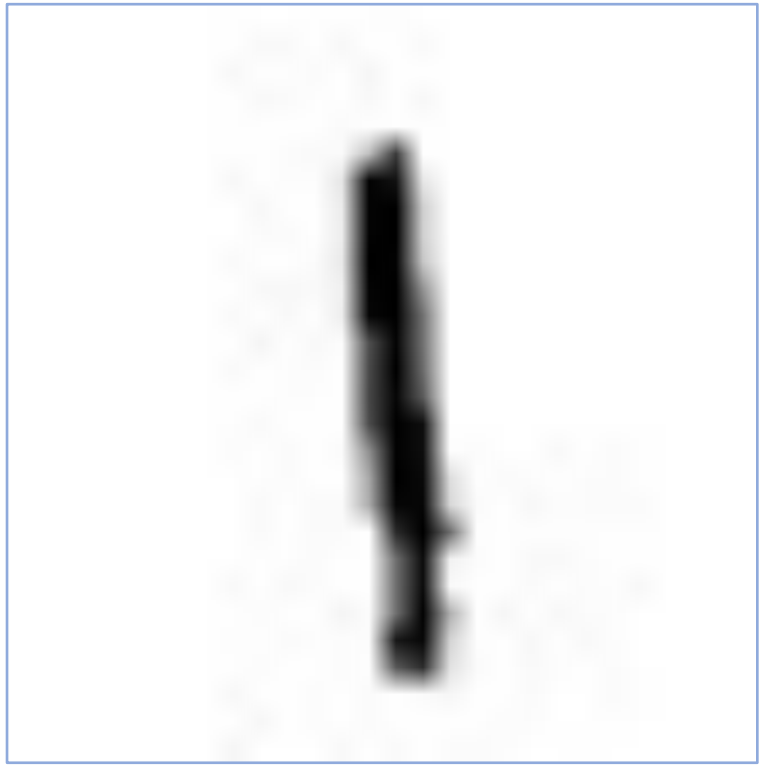
Contents

- Manifold of media data
- Reconstruction
- Localization
- Mitigation

Manifold



Manifold



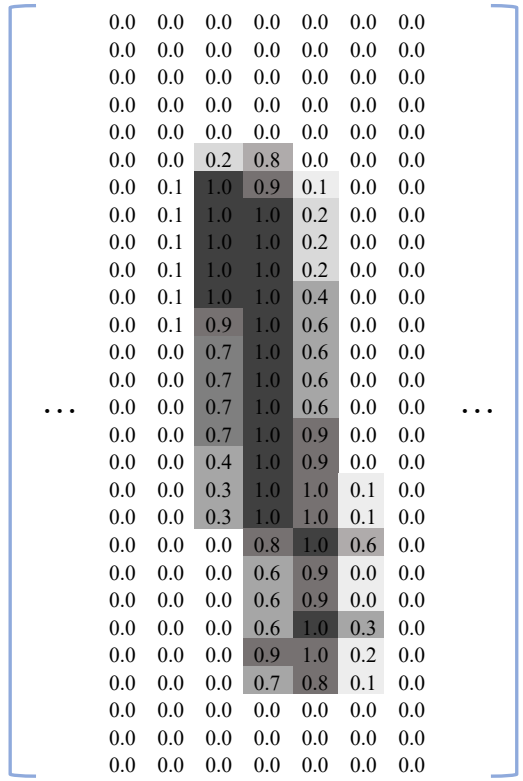
size: 1 x 30 x 30

An image in the pixel space

- A 900-dimension vector x
- Each $x[i] \in [0, 255]$

Too many dimensions!

Manifold



pixel value / 255

Not all $x \in [0, 255]^{900}$ are meaningful images; "images" of random pixel values are mostly meaningless.

- Meaningless image → ~~privacy~~
- "Perceptual" constraints over pixel values → primarily scope the privacy

Manifold

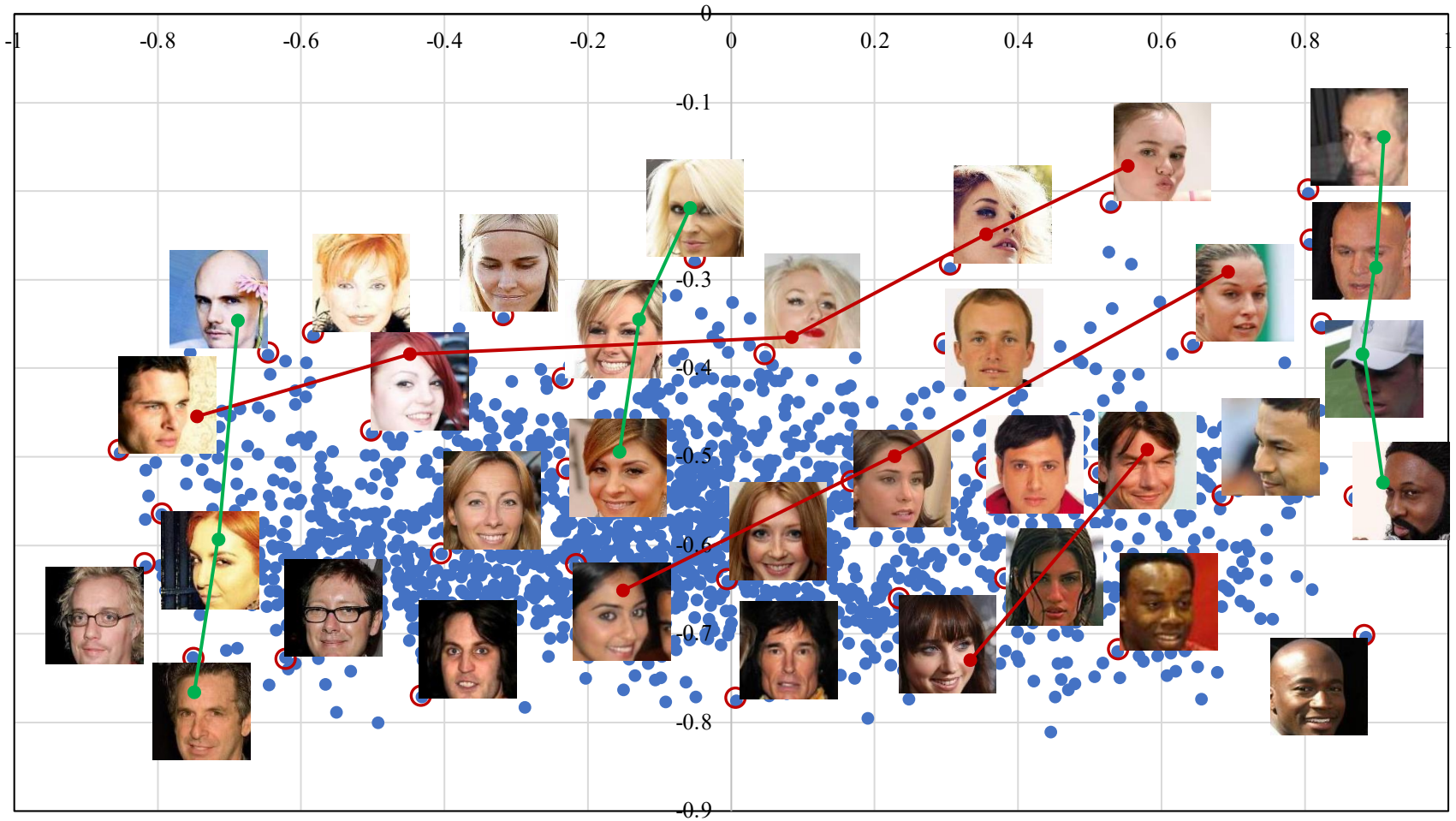
An intuitive example

Imagine that we simplify the digit “1” as a segment. Then project it onto the polar coordinate.

Only two dimensions!

Manifold

Project face photos onto a 2-dimensional manifold using our framework



Two dimensions are correlated to face colors and orientations

100 dimensions for face photos in practice

🤔

Reconstruction

side channel domain

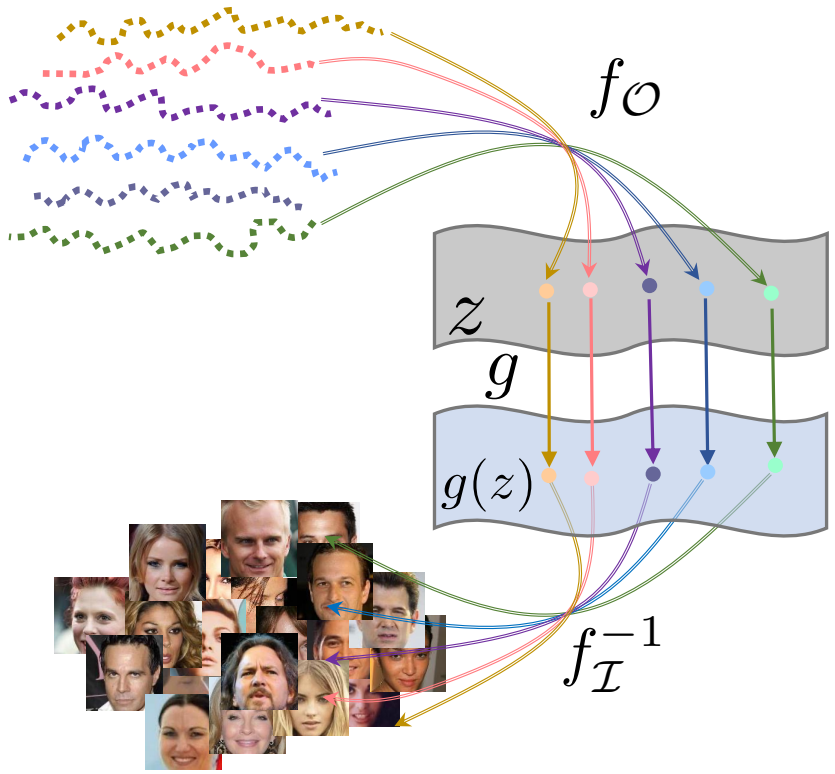


image domain

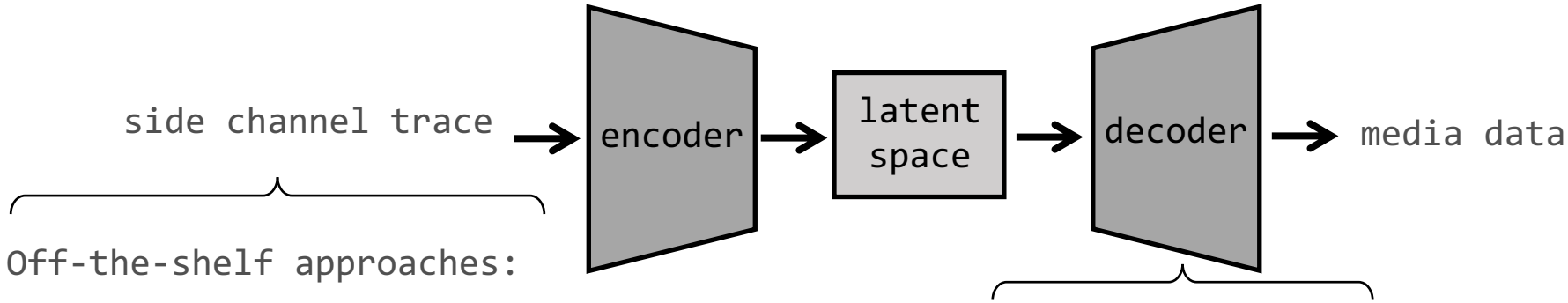
A manifold view on side channel analysis of media software

~~Data bytes (e.g., pixel values)~~ ✗

Perceptual contents (e.g., facial attributes) of much lower dimensions ✓

Reconstruction

The high-level framework



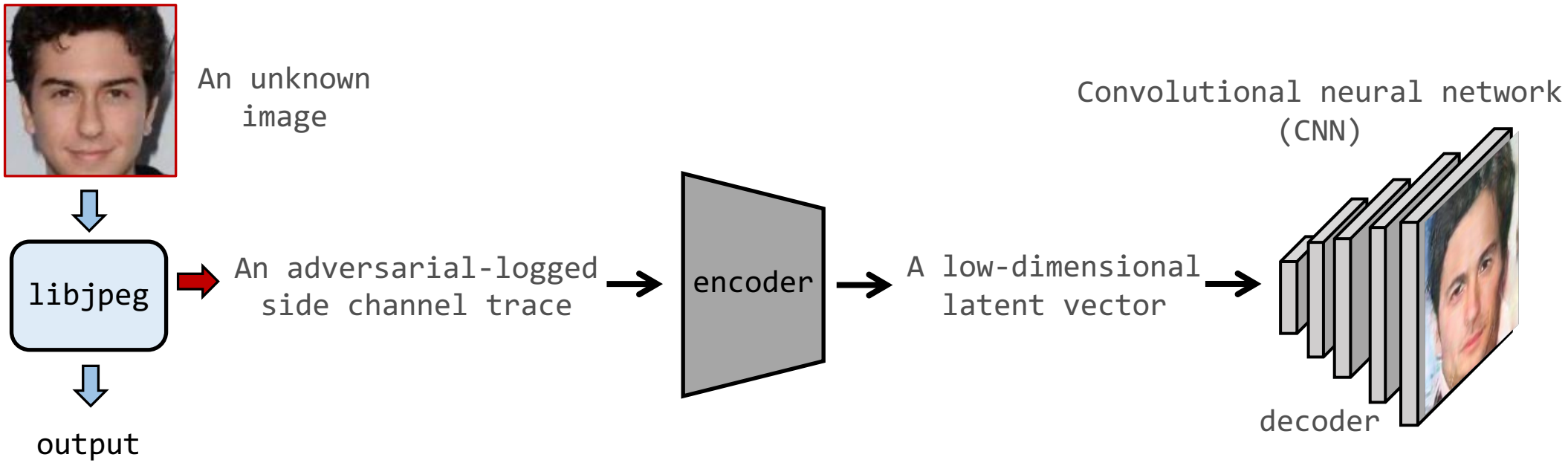
Off-the-shelf approaches:

- Prime & Probe (real attack)
- Intel Pin (“debug”)

Different decoders:

- Image
- Audio
- text

Reconstruction

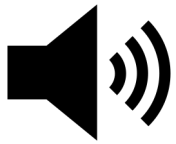


Images are continuous

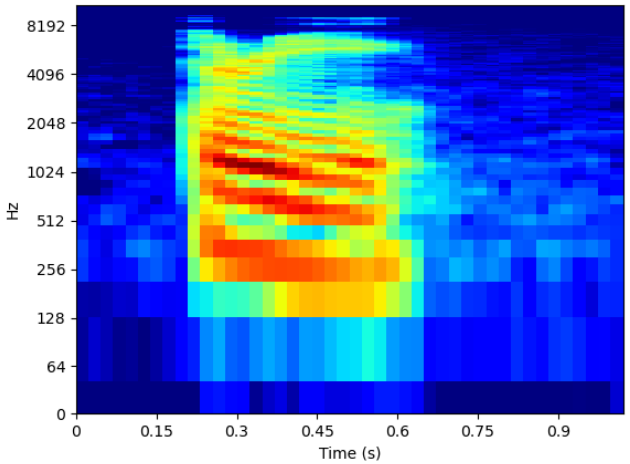
Reconstruction

raw audios

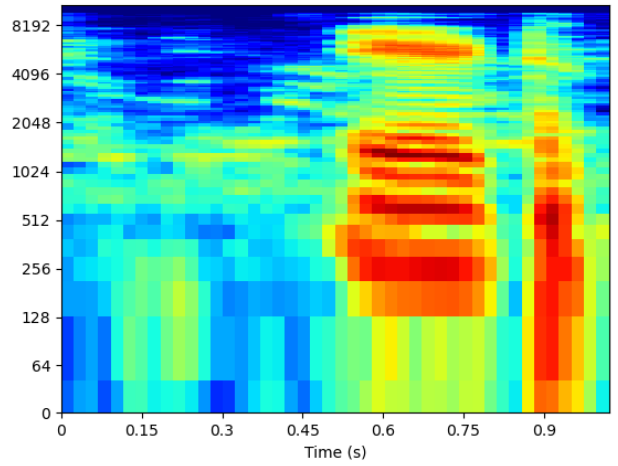
(e.g., *.wav*)



lossless

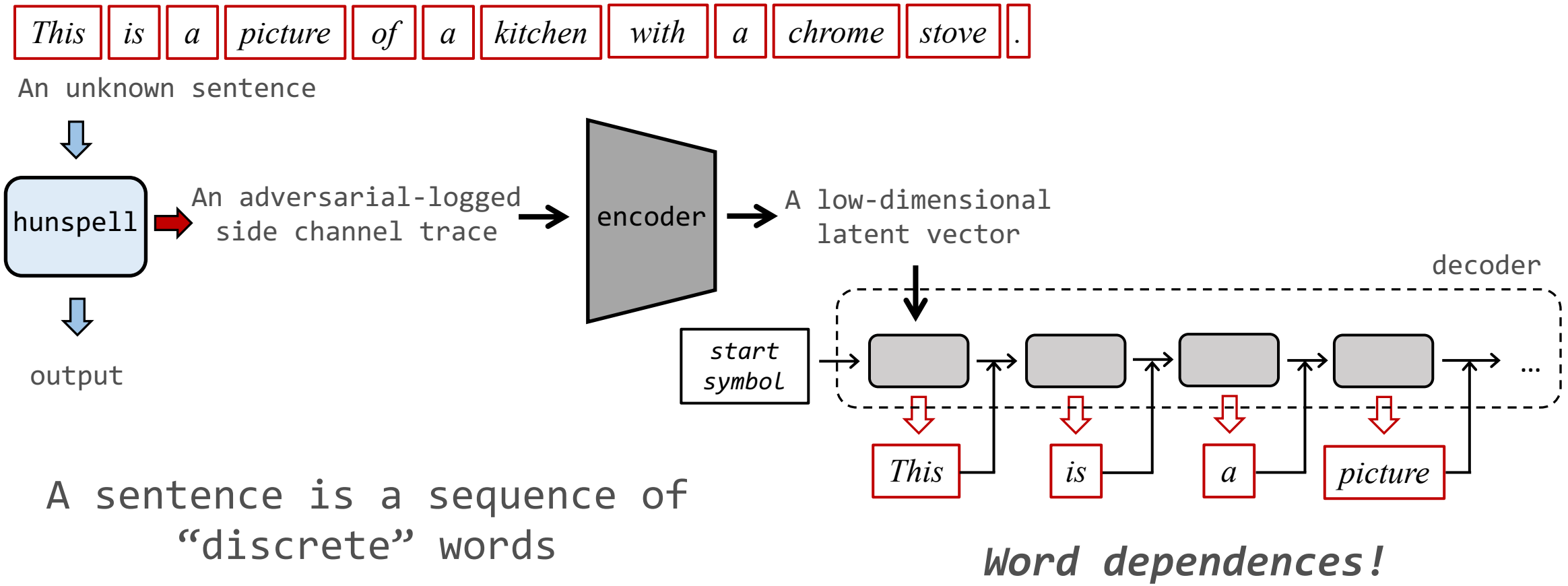


An “image”-representation of audios



log-amplitude of Mel spectrum

Reconstruction



A sentence is a sequence of "discrete" words

Word dependences!

Reconstruction

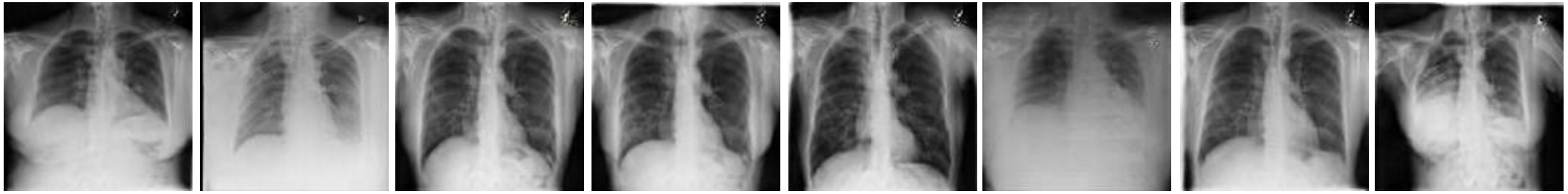
Reconstructed images

Private inputs

Reconstructed images

Private inputs

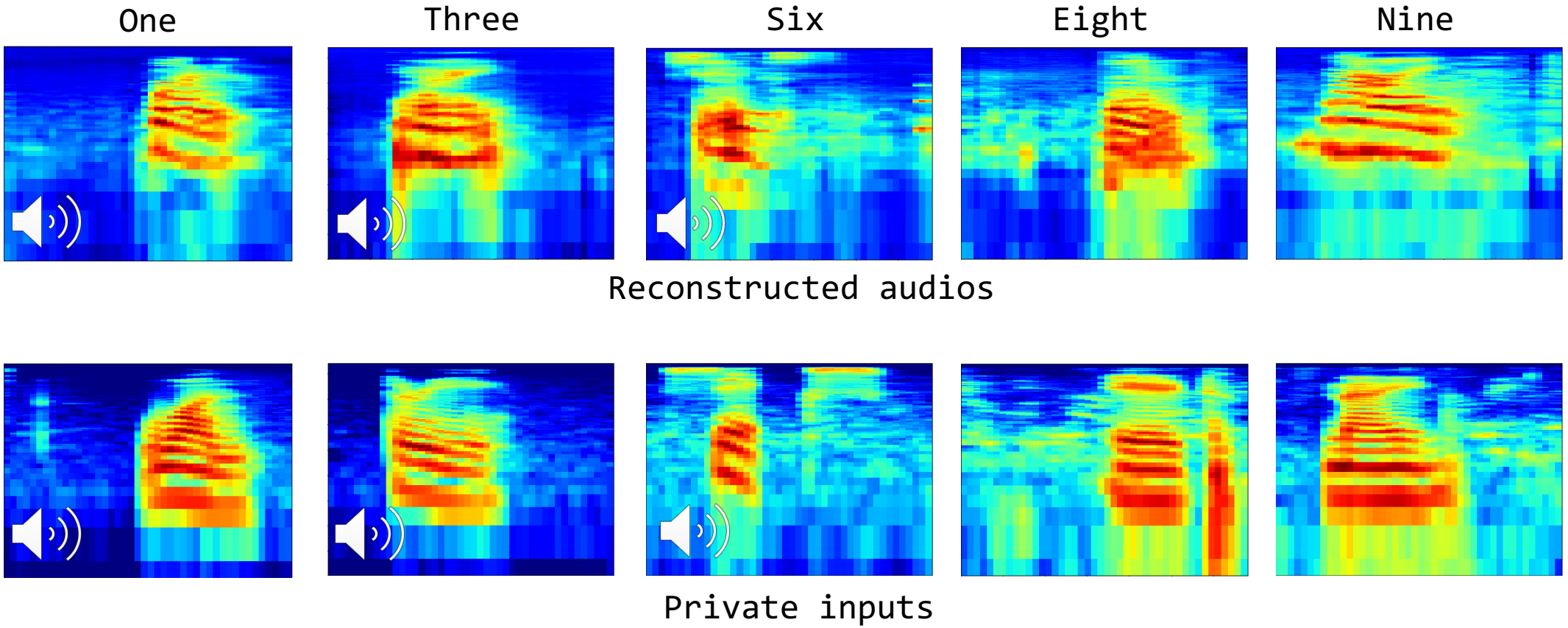
Reconstruction



Reconstructed images

Private inputs

Reconstruction

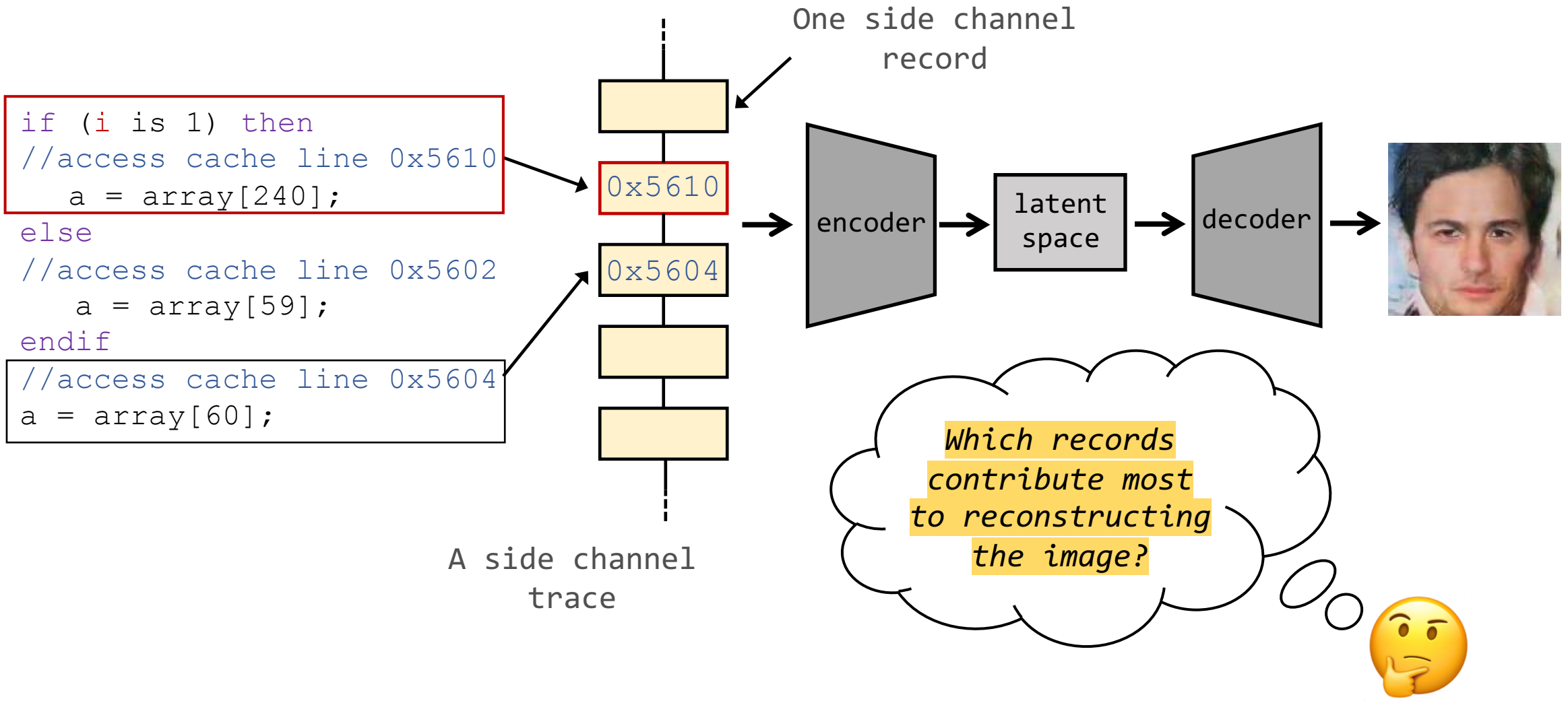


Reconstruction

Reconstructed Text	Private Input
*<UNK> I supposed to do **now** ?*	*What am I supposed to do then ?*
***I have** , the sunshine and **beautiful up** me **to the** honeymoon . The island , the sound of the <UNK> , the salty **style** air and the sunshine . . .*	*You know , the sunshine and wind remind me of our honeymoon . The island , the sound of the waves , the salty sea air and the sunshine . . .*
Mam , another minute , could I ?	*Mam , another minute , could I ?*
***It ' s** like a good idea .*	*That sounds like a good idea .*
*I <UNK>' t want to insult Jill or her **brother** . I think Jill **,** could be it . But I ' **ll** rather have some **to** little older .*	*I don ' t want to insult Jill or her mother . I think Jill maybe could do it . But I ' d rather have someone a little older .*
*I think it **'** be better **for find** a good babysitter here . It **' be cost , an** or three days .*	*I think it would be better to have a good babysitter here . It might even be for two or three days .*
*She **is** a **single** cold , and **it** don ' t want to take **care to** us . But we don ' t know **how** can stay with her .*	*She has a bad cold , and we don ' t want to take her with us . But we don ' t know who can stay with her .*
*This is **very** <UNK>, I **have** . But Hank and I are leaving tonight .*	*This is short notice , I know . But Hank and I are leaving tonight .*
*I ' m sorry **, say that** . What ' s wrong with her ?*	*I ' m sorry to hear it . What ' s wrong with her ?*
*Have you ever **been** a parking ticket ?*	*Have you ever gotten a parking ticket ?*

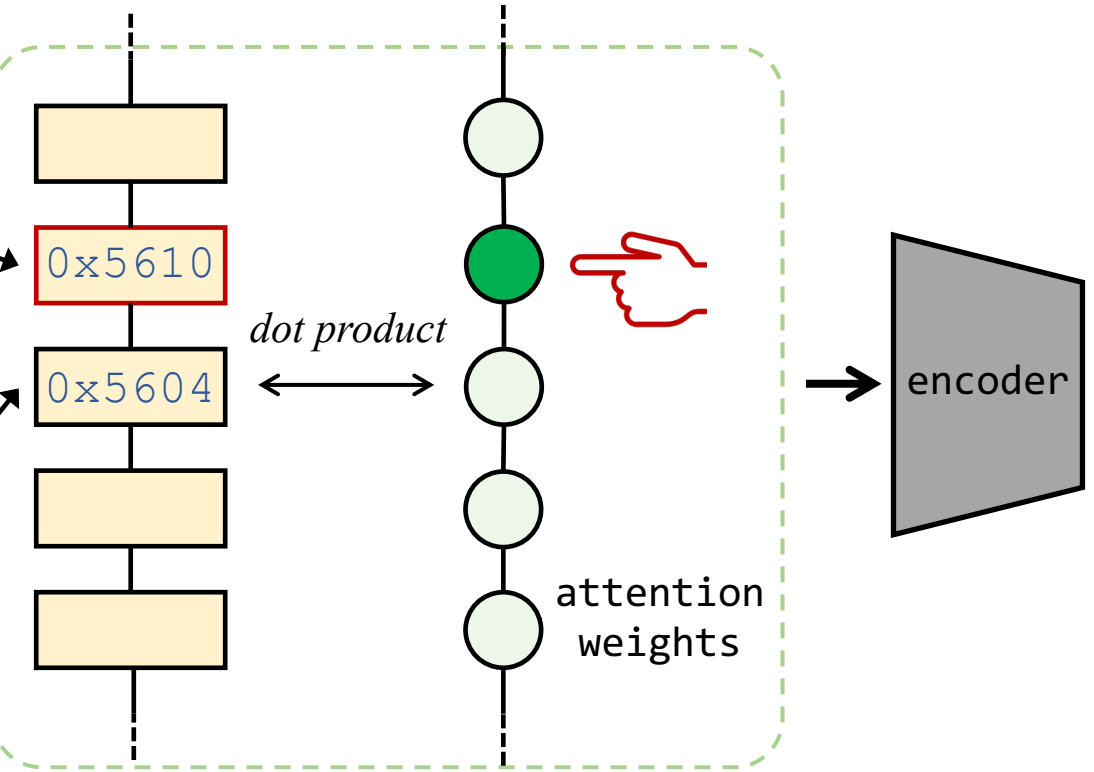
Localization

```
if (i is 1) then
//access cache line 0x5610
   a = array[240];
else
//access cache line 0x5602
   a = array[59];
endif
//access cache line 0x5604
a = array[60];
```



Localization

```
if (i is 1) then
//access cache line 0x5610
   a = array[240];
else
//access cache line 0x5602
   a = array[59];
endif
//access cache line 0x5604
a = array[60];
```



Neural attention! 😎

Localization

```
1  int HUFF_EXTEND(int x, int s) {
2    // ‘‘ex_test’’ and ‘‘ex_offset’’ are
3    // pre-calculated arrays
4    if (x < ex_test[s])
5      return x + ex_offset[s];
6    else
7      return x;
8    }
9
10 boolean decode_mcu_fast(j_decompress_ptr cinfo,
11   JBLOCKROW *MCU_data) {
12   huff_entropy_ptr entropy =
13    (huff_entropy_ptr)cinfo->entropy;
14   /* preprocessing */
15   for (int i = 0; i < cinfo->blocks_in_MCU; i++)
16     d_derived_tbl *dctbl = entropy->dc_cur_tbls[i];
17     int s, k, r, l;
18     /* get index ‘‘idx’’ based on ‘‘s’’ */
19     /* update ‘‘r’’ */
20     s = dctbl->lookup[idx];
21     // ‘‘lookup’’ is pre-calculated array
22     if (s)
23       s = HUFF_EXTEND(r, s);
24     }
25     /* do something */
26   }
27   /* do something and return */
28 }
```

Localized vulnerabilities in libjpeg

- Minimum coded unit (MCU)-related modules
- Inverse discrete cosine transformation (IDCT)-related modules [1] [2]
- Other image transformation routines and output dumping routines

[1] *Controlled-channel attacks: Deterministic side channels for untrusted operating systems.* In S&P 2015.
[2] *High-resolution side channels for untrusted operating systems.* In USENIX ATC 2017.

Mitigation

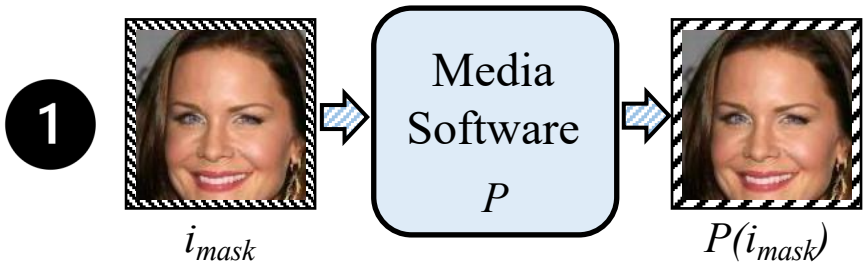
Media software: process **data bytes**
(e.g., a pixel value)

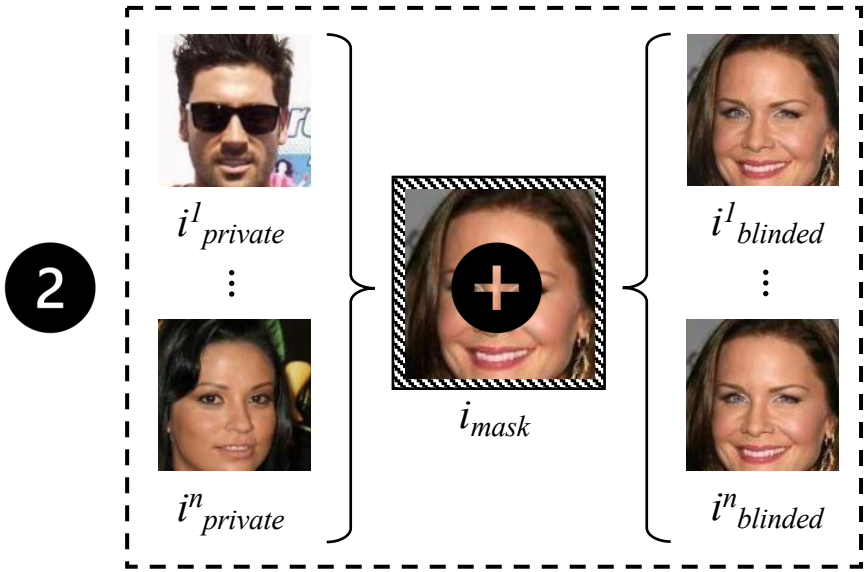
V.S.

Our attack: focus on **perceptions**
(e.g., facial attributes)

"Blind" the perceptions!

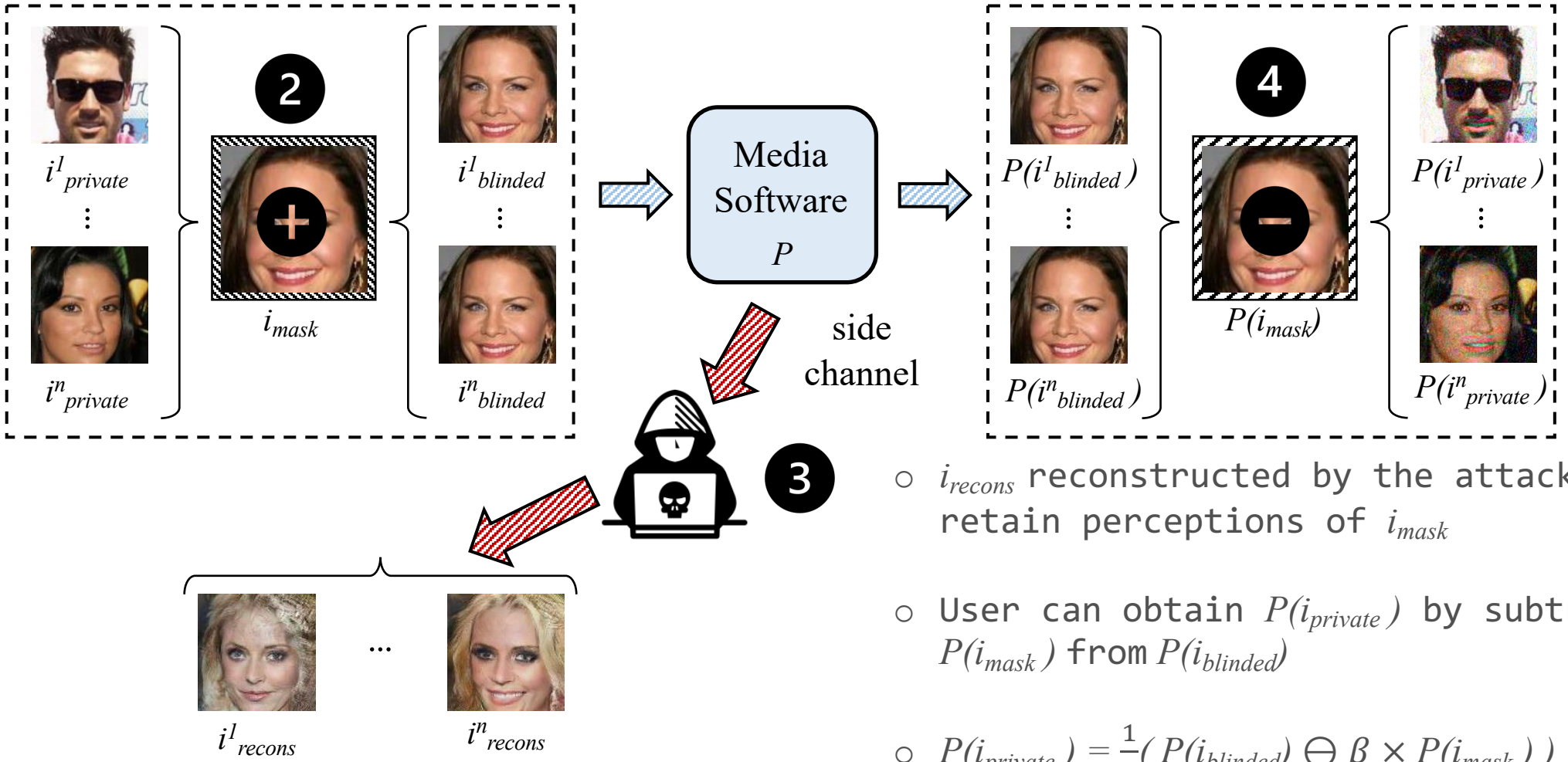
Mitigation





- Randomly pick one universal mask i_{mask}
- Pre-compute $P(i_{mask})$
- i_{mask} must be perceptually correlated to private input $i_{private}$ (e.g., both are face photos)
- Set $i_{blinded} = \alpha \times i_{private} \oplus \beta \times i_{mask}$, rather than $i_{private}$, as the input of P
- $\beta \gg \alpha$ and $\alpha + \beta = 1$

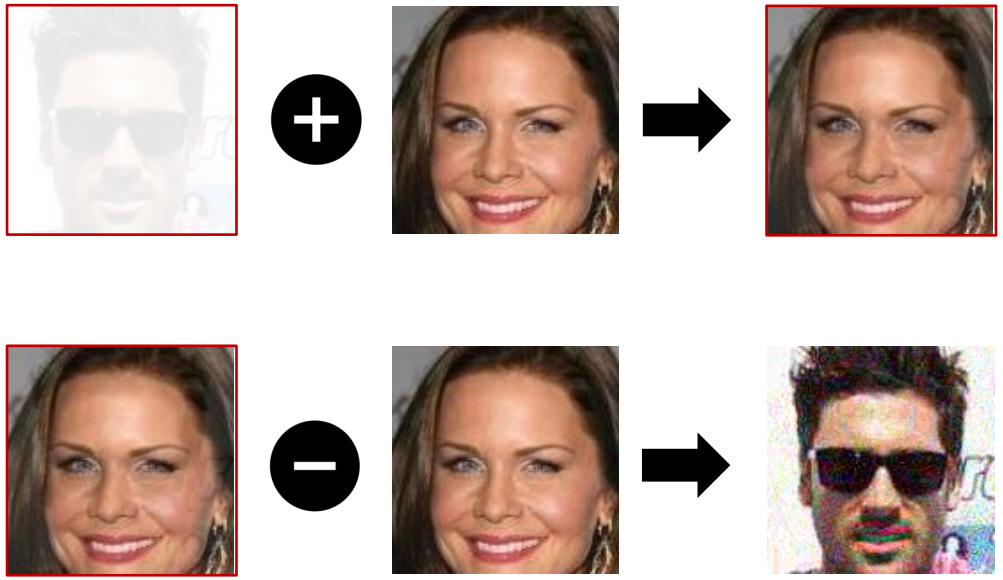
Mitigation



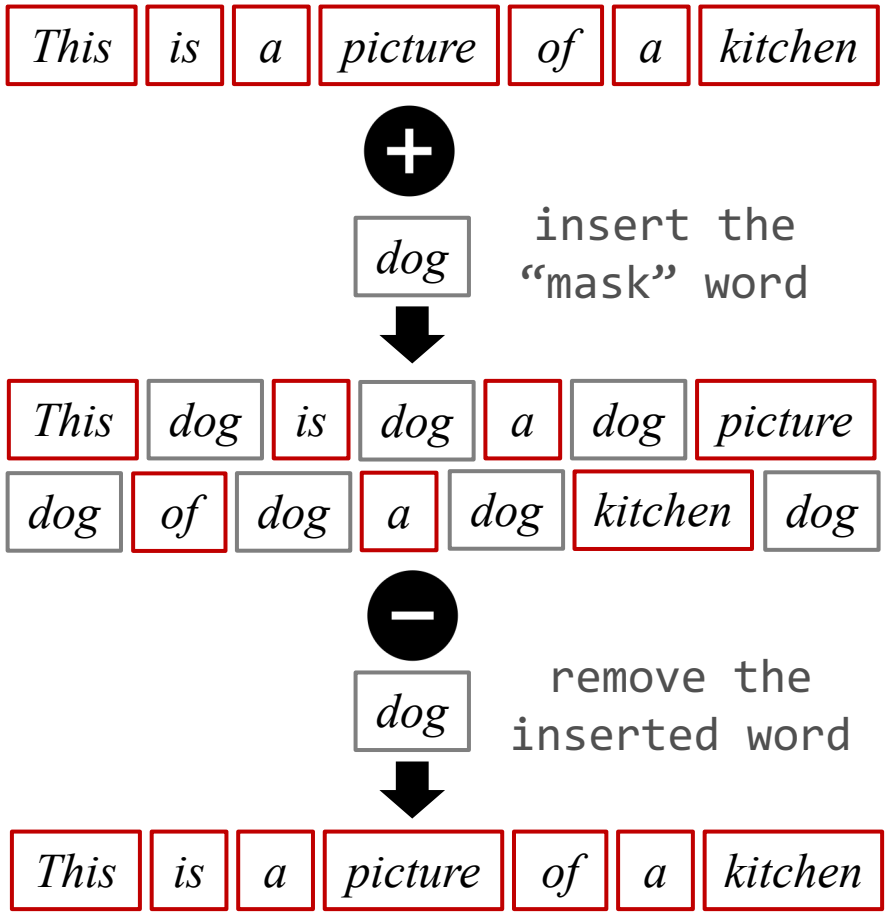
- i_{recons} reconstructed by the attacker mostly retain perceptions of i_{mask}
- User can obtain $P(i_{private})$ by subtracting $P(i_{mask})$ from $P(i_{blinded})$
- $P(i_{private}) = \frac{1}{\alpha}(P(i_{blinded}) \ominus \beta \times P(i_{mask}))$

Mitigation

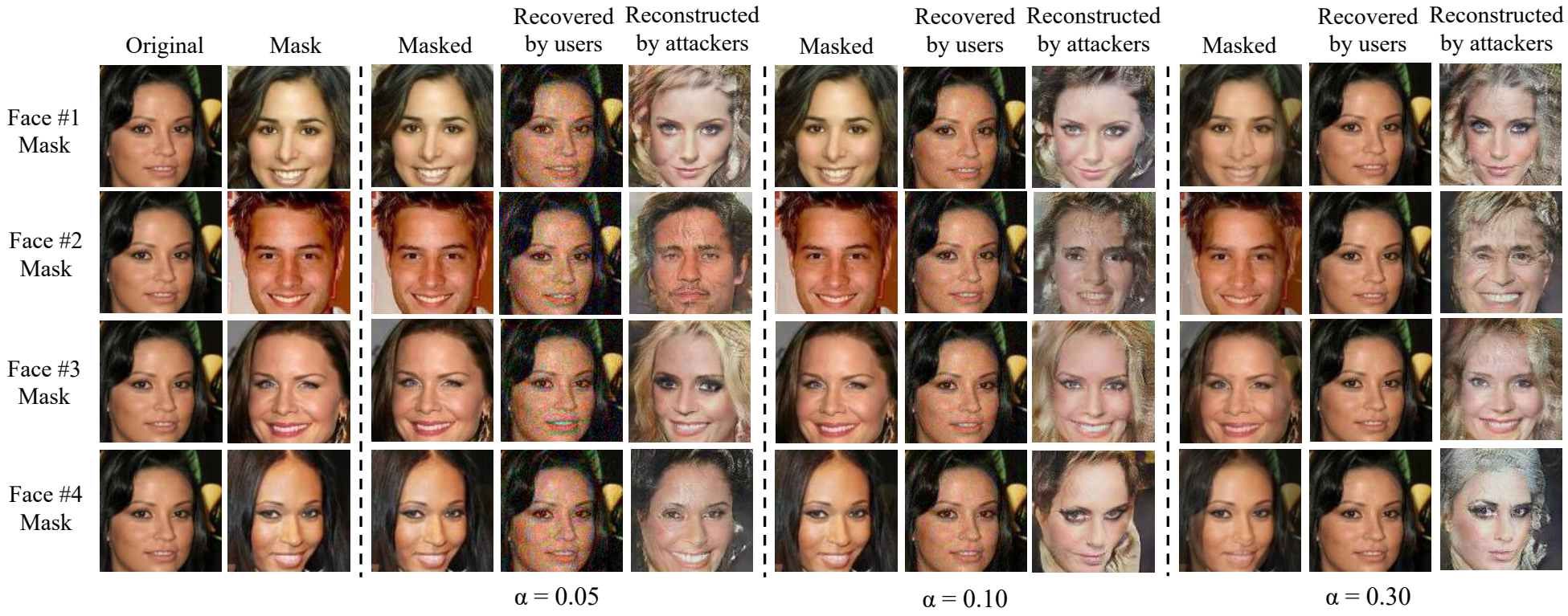
lower weight higher weight



data bytes addition & subtraction



Mitigation



- Attacker only reconstructs perceptions of the mask
- User can recover the perceptions of private inputs

Thank you for listening!

Contact Yuanyuan Yuan (https://yuanyuan-yuan.github.io/)
for more details

https://arxiv.org/pdf/2112.04947.pdf

Preprint

(an extended version of 35 pages)

https://github.com/Yuanyuan-Yuan/Manifold-SCA

Artifact

香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

usenix
THE ADVANCED
COMPUTING SYSTEMS
ASSOCIATION