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
Threat Model & Overview

Have you ever gotten a parking ticket?

- Media Software
  - image
  - audio
  - text

Private media input

Cache access
OS page table access

Have you ever been a parking ticket?

reconstructed media input

- 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

What is “Manifold”?  
Dimension reduction!
Manifold

An image in the pixel space

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

Too many dimensions!
Manifold

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

- Meaningless image $\rightarrow$ privacy
- “Perceptual” constraints over pixel values $\rightarrow$ 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!
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

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

- Encoder
  - Latent space
  - Decoder
  - Media data

Side channel trace

Off-the-shelf approaches:
- Prime & Probe (real attack)
- Intel Pin ("debug")

Different decoders:
- Image
- Audio
- Text
Reconstruction

An adversarial-logged side channel trace → encoder → A low-dimensional latent vector → decoder

Convolutional neural network (CNN)

Images are continuous
Reconstruction

An “image”-representation of audios

raw audios (e.g., .wav)

log-amplitude of Mel spectrum
Reconstruction

This is a picture of a kitchen with a chrome stove.

An unknown sentence

hunspell

An adversarial-logged side channel trace

encoder

A low-dimensional latent vector

decoder

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

Private inputs

Reconstructed audios
<table>
<thead>
<tr>
<th>Reconstructed Text</th>
<th>Private Input</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;UNK&gt; I supposed to do now ?</td>
<td>What am I supposed to do then ?</td>
</tr>
<tr>
<td>I have , the sunshine and beautiful up me to the honeymoon . The island , the sound of the &lt;UNK&gt; , the salty style air and the sunshine . . .</td>
<td>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 . . .</td>
</tr>
<tr>
<td>Mam , another minute , could I ?</td>
<td>Mam , another minute , could I ?</td>
</tr>
<tr>
<td>It ' s like a good idea .</td>
<td>That sounds like a good idea .</td>
</tr>
<tr>
<td>I &lt;UNK&gt;'t want to insult Jill or her brother . I think Jill could be it . But I 'll rather have some to little older .</td>
<td>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 .</td>
</tr>
<tr>
<td>I think it ' be better for find a good babysitter here . It ' be cost , an or three days .</td>
<td>I think it would be better to have a good babysitter here . It might even be for two or three days .</td>
</tr>
<tr>
<td>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 .</td>
<td>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 .</td>
</tr>
<tr>
<td>This is very &lt;UNK&gt; , I have . But Hank and I are leaving tonight .</td>
<td>This is short notice , I know . But Hank and I are leaving tonight .</td>
</tr>
<tr>
<td>I ' m sorry , say that . What ' s wrong with her ?</td>
<td>I ' m sorry to hear it . What ' s wrong with her ?</td>
</tr>
<tr>
<td>Have you ever been a parking ticket ?</td>
<td>Have you ever gotten a parking ticket ?</td>
</tr>
</tbody>
</table>
Localization

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

A side channel trace

One side channel record

encoder \rightarrow \text{latent space} \rightarrow \text{decoder}

Which records contribute most to reconstructing the image?
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

```c
int HUFF_EXTEND(int x, int s) {
    // 'ex_test' and 'ex_offset' are pre-calculated arrays
    if (x < ex_test[s])
        return x + ex_offset[s];
    else
        return x;
}
```

boolean decode_mcu_fast(j_decompress_ptr cinfo, JBLOCKROW *MCU_data) {
    huff_entropy_ptr entropy = (huff_entropy_ptr)cinfo->entropy;
    /* preprocessing */
    for (int i = 0; i < cinfo->blocks_in_MCU; i++)
        d_derived_tbl *dctbl = entropy->dc_cur_tbls[i];
    int s, k, r, l;
    /* get index 'idx' based on 's' */
    /* update 'r' */
    s = dctbl->lookup[idx];
    // ''lookup'' is pre-calculated array
    if (s)
        s = HUFF_EXTEND(r, s);
    /* do something */
}
/* do something and return */
```

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

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

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

\[ P(\text{i}_{\text{private}}^{\text{1}}) = \frac{1}{\alpha} ( P(\text{i}_{\text{blinded}}^{\text{1}}) \ominus \beta \times P(\text{i}_{\text{mask}}^{\text{1}}) ) \]

- \( \text{i}_{\text{recons}} \) reconstructed by the attacker mostly retain perceptions of \( \text{i}_{\text{mask}} \)
- User can obtain \( P(\text{i}_{\text{private}}^{\text{1}}) \) by subtracting \( P(\text{i}_{\text{mask}}^{\text{1}}) \) from \( P(\text{i}_{\text{blinded}}^{\text{1}}) \)
Mitigation

lower weight  higher weight

data bytes addition & subtraction

This is a picture of a kitchen

+ insert the “mask” word

dog

This dog is dog a dog picture
dog of dog a dog kitchen dog

- remove the inserted word

dog

This is a picture of a kitchen
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

Preprint
(an extended version of 35 pages)

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