UniID: Spoofing Face Authentication By Universal Identity

Zhihao Wu[1], Yushi Cheng*[1, 2], Shibo Zhang[1], Xiaoyu Ji*[1], Wenyuan Xu[1]

[1] Ubiquitous System Security Lab (USSLAB), Zhejiang University

[2] ZJU-UIUC Institute, Zhejiang University

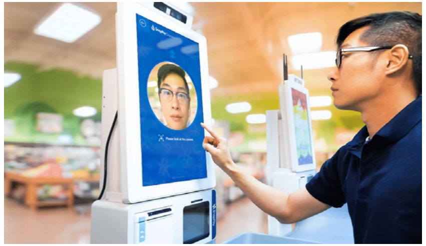
{zhihaowu, yushicheng, zhsb, xji, wyxu}@zju.edu.cn

Face Authentication Systems are everywhere!

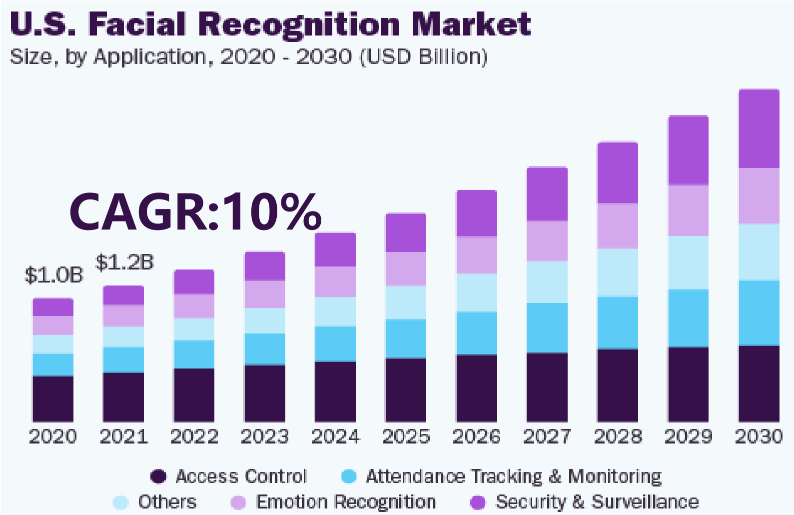
Smart Phone Unlock

Access Control

Home Unlock



Financial Payments



Source: Grandviewresearch

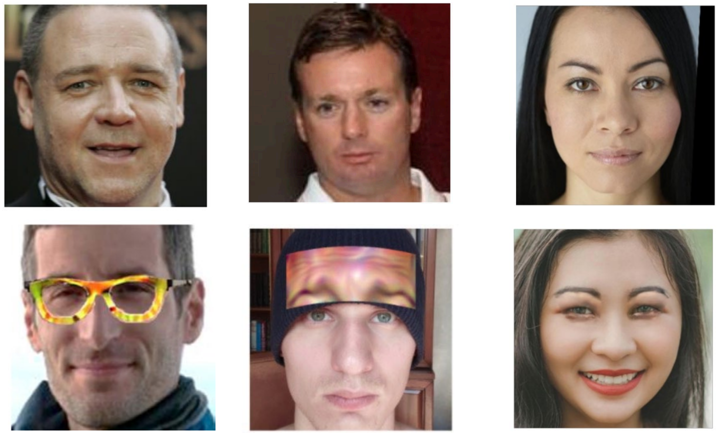
Are face authentication systems secure?

Spoofing Face Authentication Systems

- ***Adversarial Attacks***

Target

Attacker



Adv-Glass Adv-Hat Adv-Makeup

Spoofing Face Authentication Systems

- ***Adversarial Attacks***

Properties：

- **Specially designed (1v1)**
- **One-time effective**
- **Easily detectable**

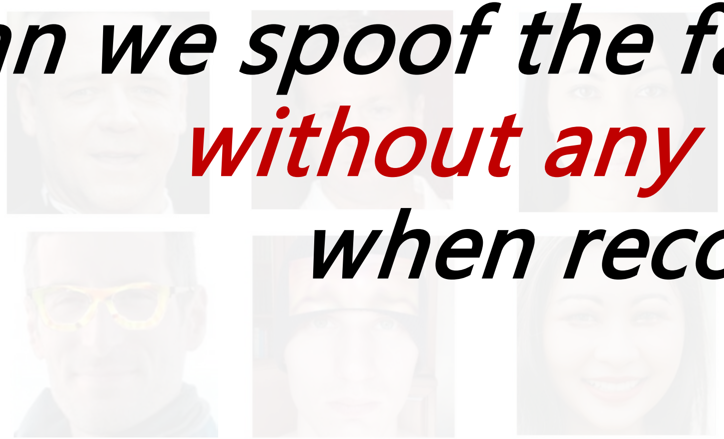
Not practical and stealthy enough in the real-world

Spoofing Face Authentication Systems

- Adversarial Attacks

Target

Attacker



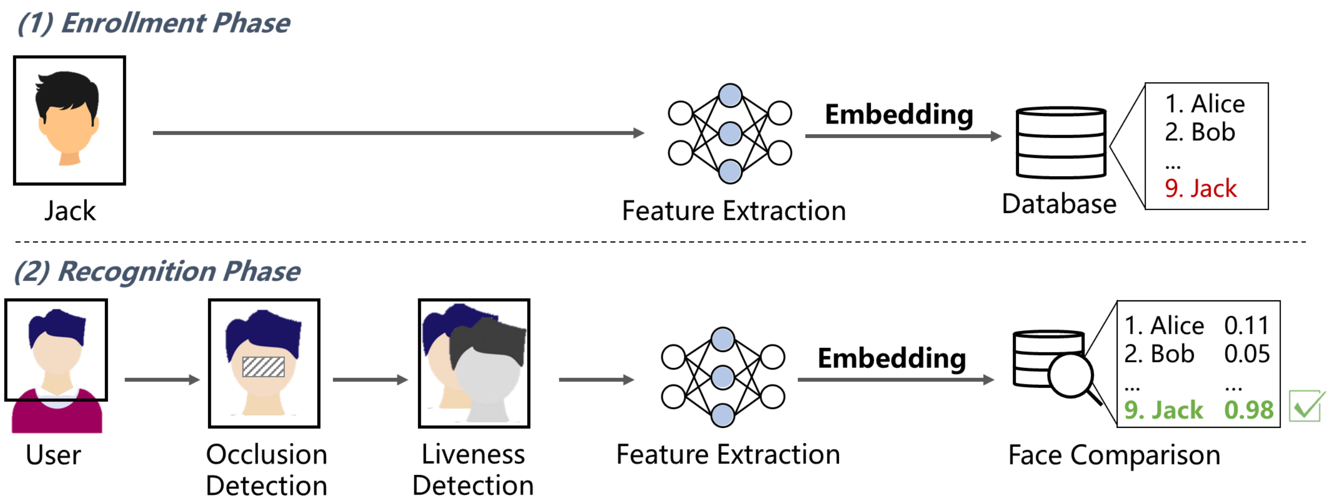
Adv-Glass | Adv-Hat | Adv-Makeup

Easily detectable

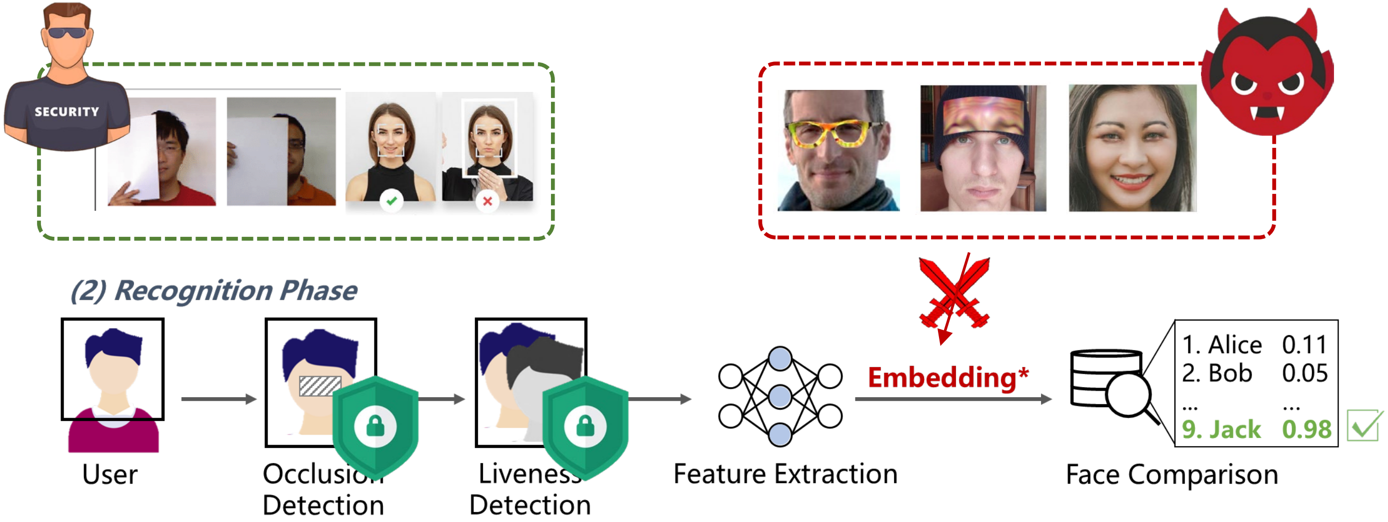
Not practical and stealthy enough in the real-world

Can we spoof the face authentication
without any camouflage
when recognizing?

Face authentication system

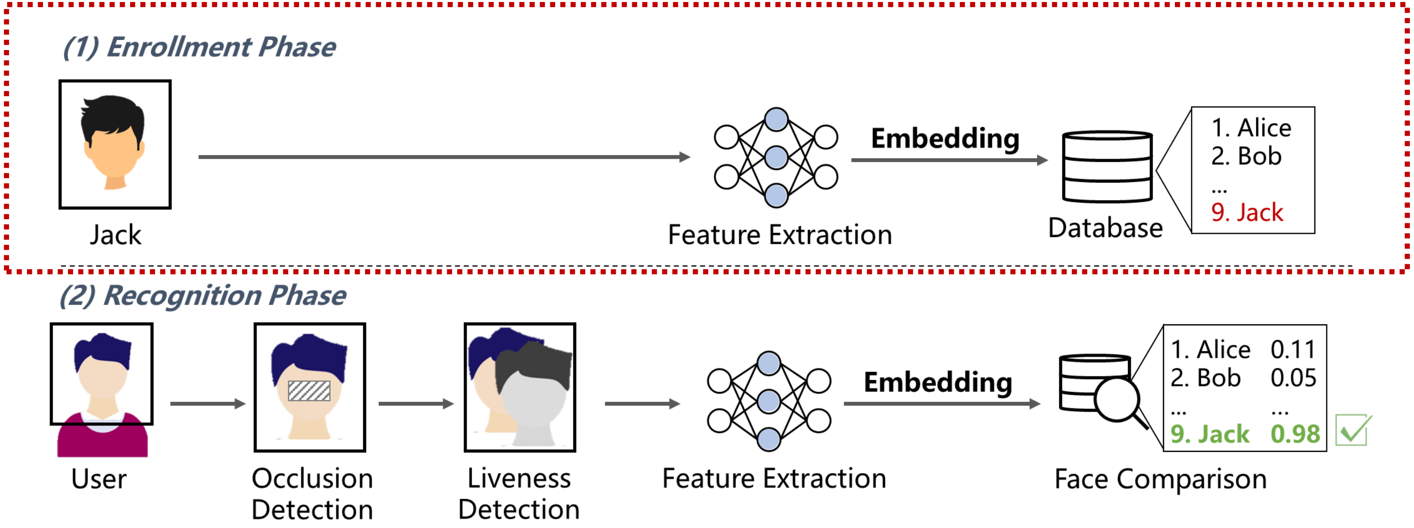


Existing attacks and defenses

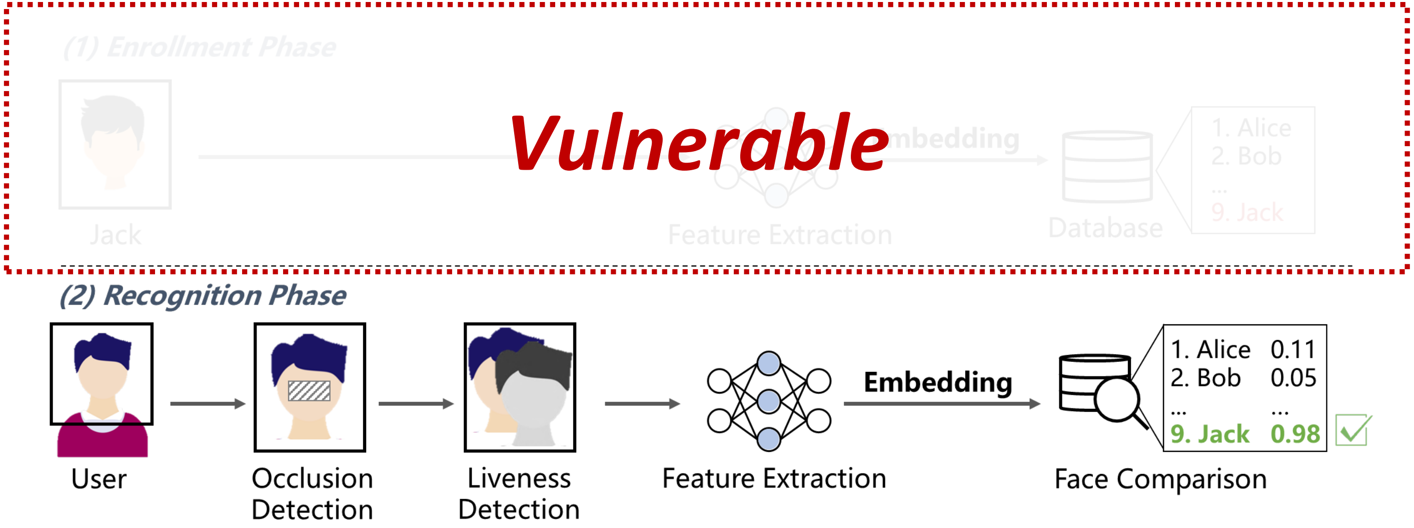


Spoof attacks through recognition phase become difficult!

The enrollment phase is overlooked!

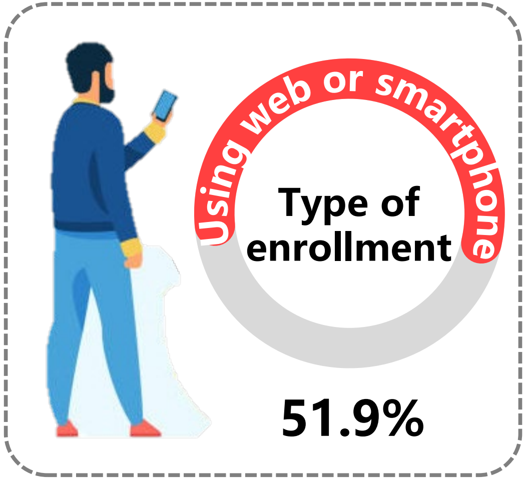


The enrollment phase is overlooked！



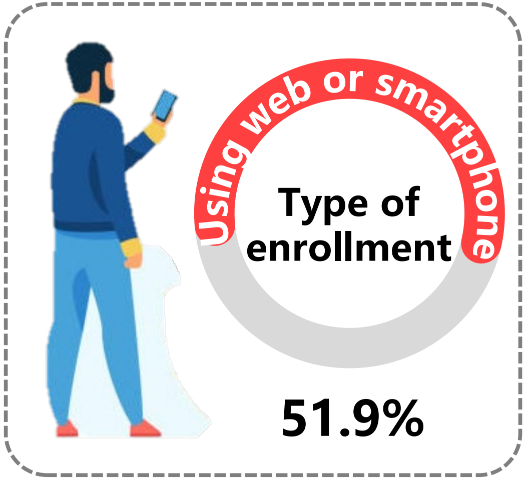
Vulnerabilities in the enrollment phase

Self-uploading



Vulnerabilities in the enrollment phase

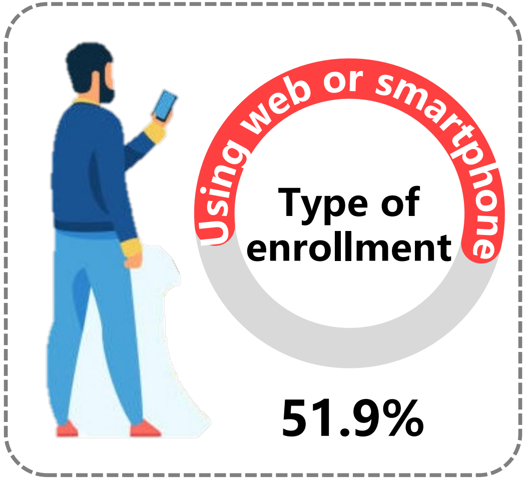
Self-uploading



Unsupervised

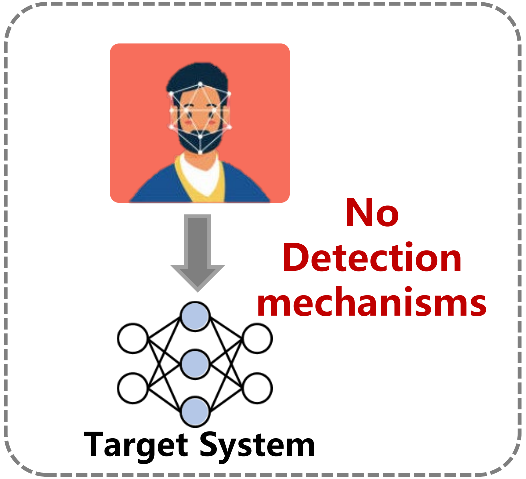
Vulnerabilities in the enrollment phase

Self-uploading



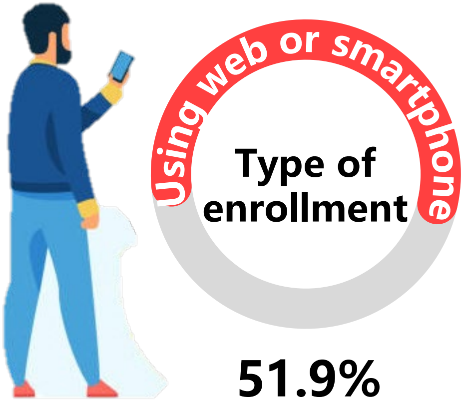
Unsupervised

Unchecked



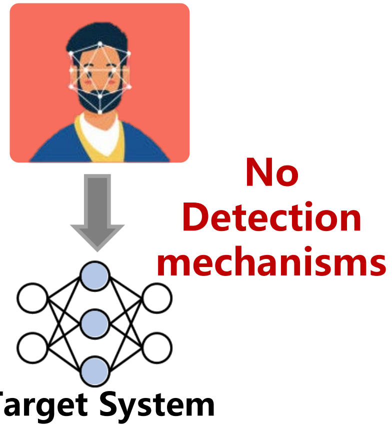
Vulnerabilities in the enrollment phase

Self-uploading



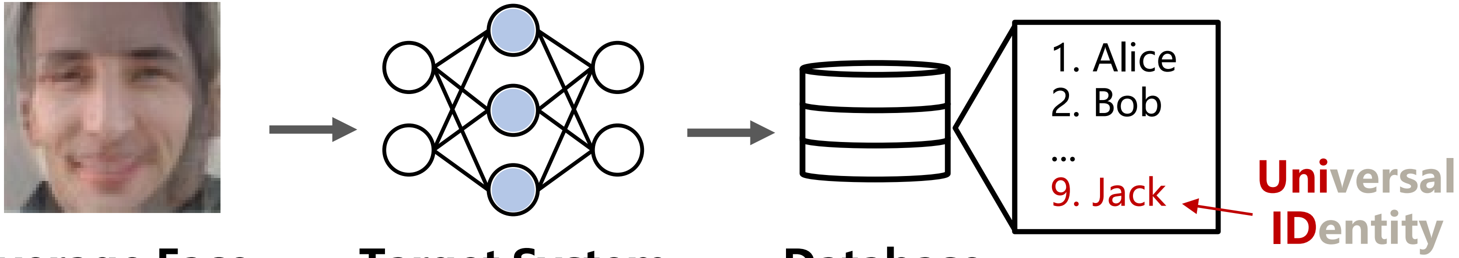
Unsupervised

Unchecked



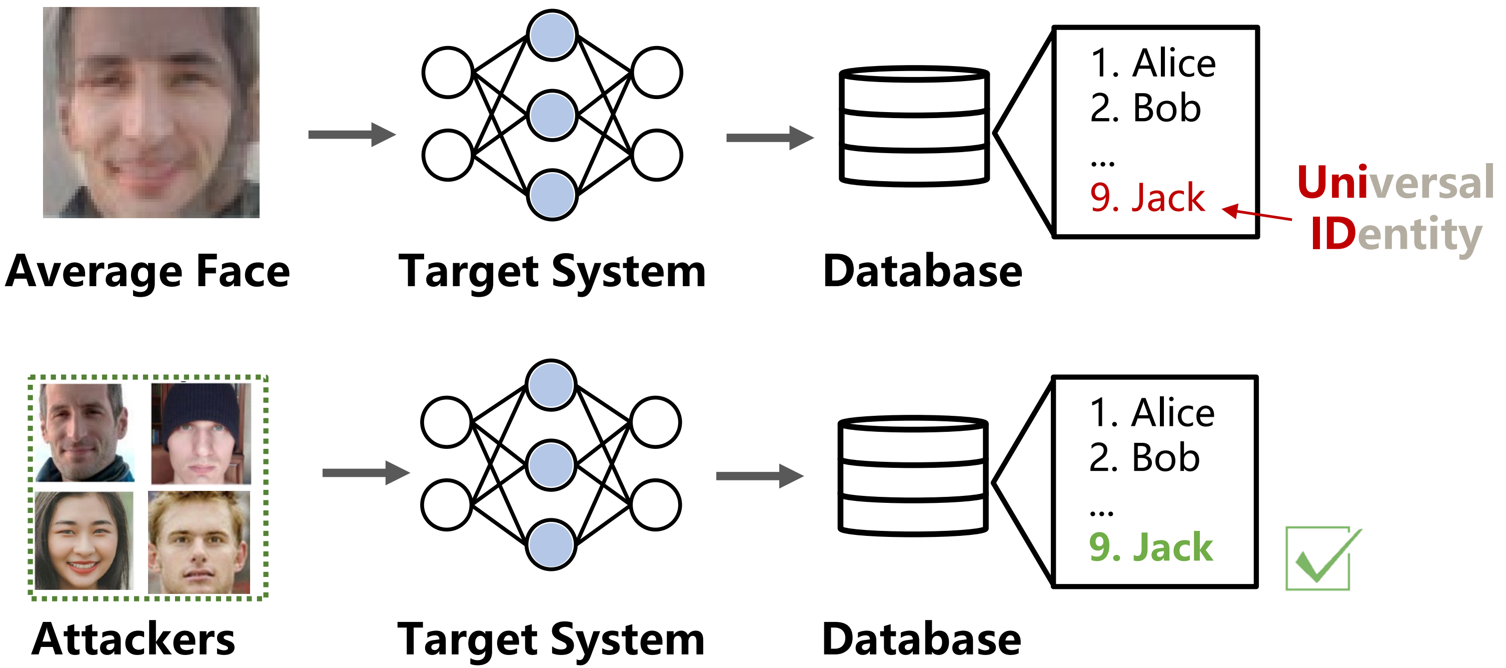
The enrollment phase can be a new entry point for spoofing attacks!

Our basic idea

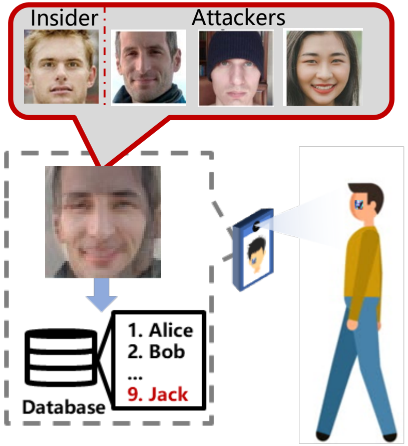


Average Face **Target System** **Database**

Our basic idea

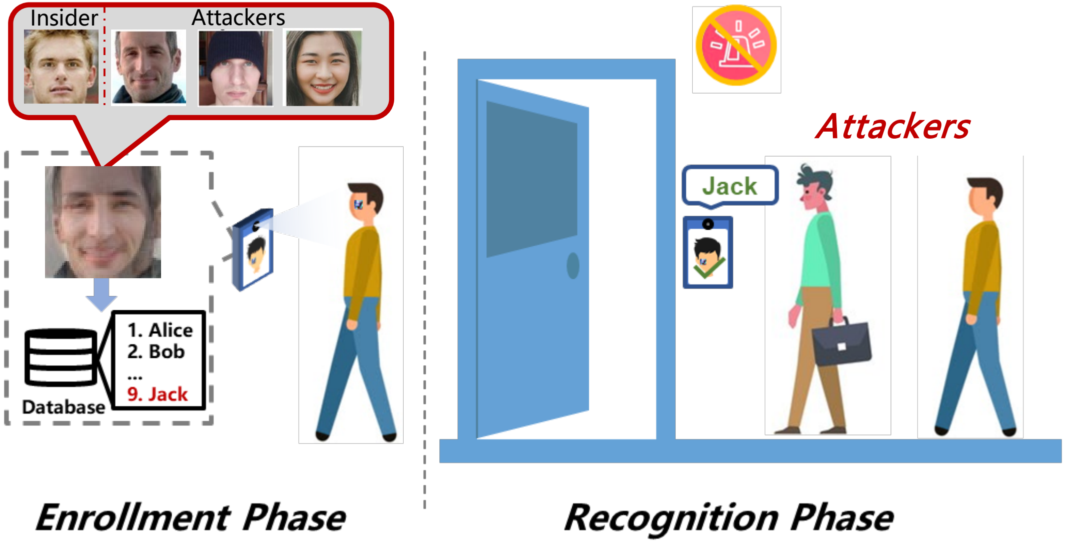


What UniID can do?

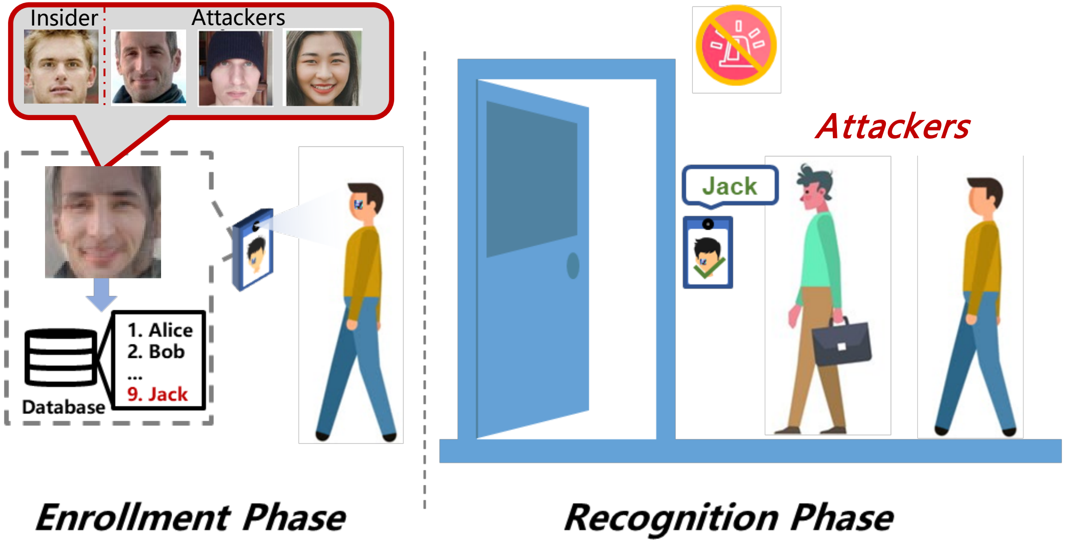


Enrollment Phase

What UniID can do?



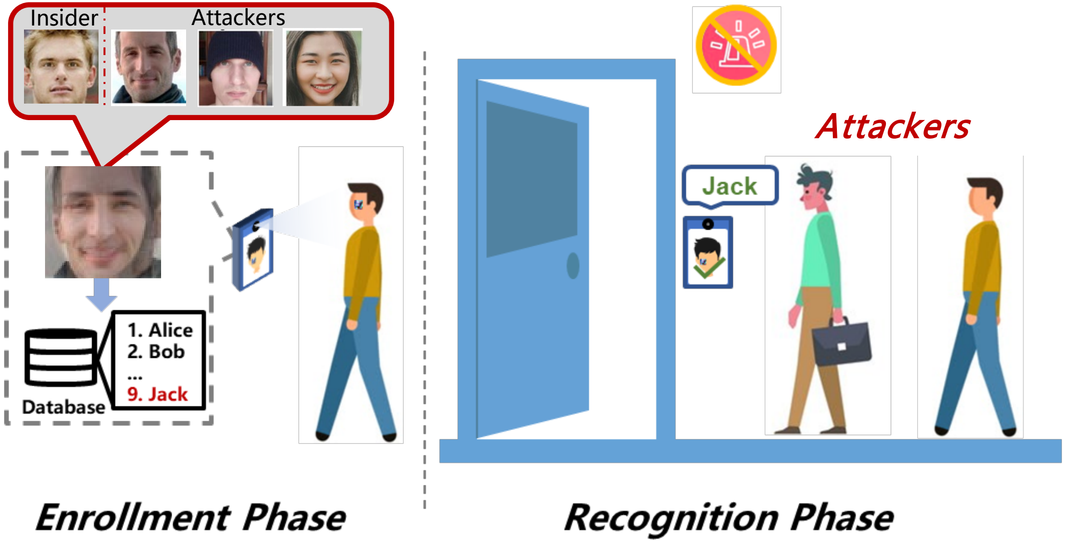
What UniID can do?



Capabilities :

- ☐ **Multiple attackers**
- ☐ **Inconspicuous**

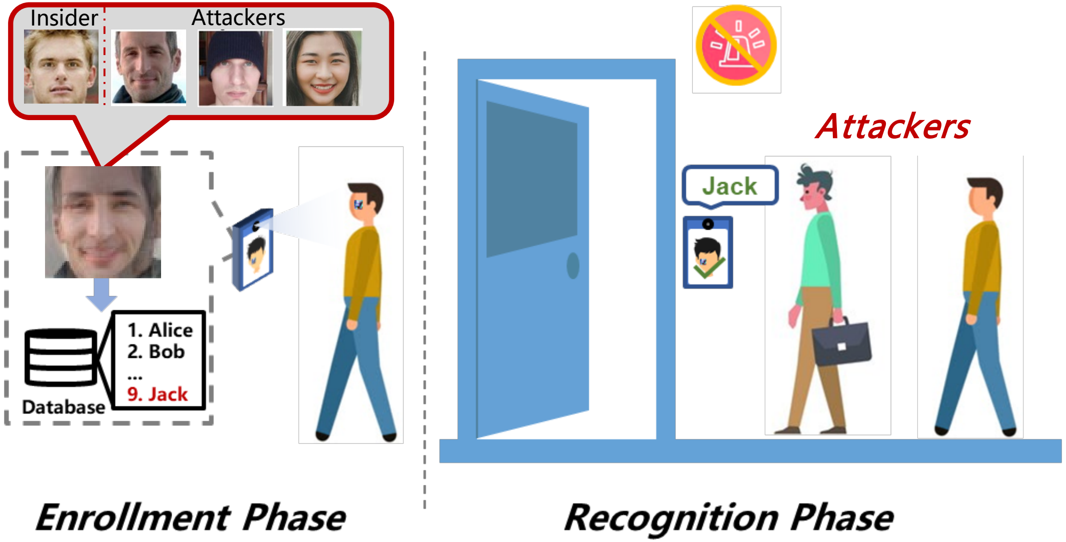
What UniID can do?



Capabilities :

- ☐ **Multiple attackers**
- ☐ **Inconspicuous**
- ☐ **One-time enrollment, unlimited-times spoofing**

What UniID can do?

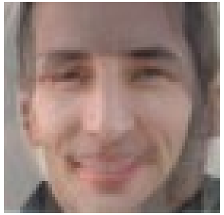


Capabilities：

- **Multiple attackers**
- **Inconspicuous**
- **One-time enrollment, unlimited-times spoofing**

Injecting UniID sounds intuitive, but not trivial.

How to achieve UniID…

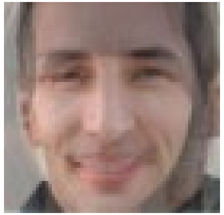


Average Face

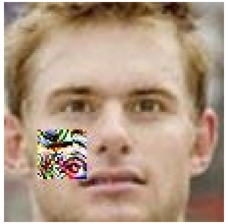
Facts：

- ***Attackers have no permission to access the database***
- ***Average face doesn't exist in real life***

How to achieve UniID…



Average Face



Insider

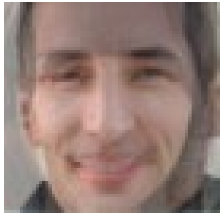
Facts：

- ***Attackers have no permission to access the database***
- ***Average face doesn't exist in real life***

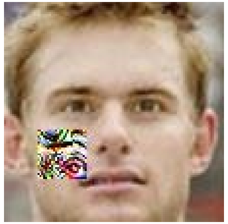
Our method:

- ***Try to find an "average face" at the feature level***

How to achieve UniID…

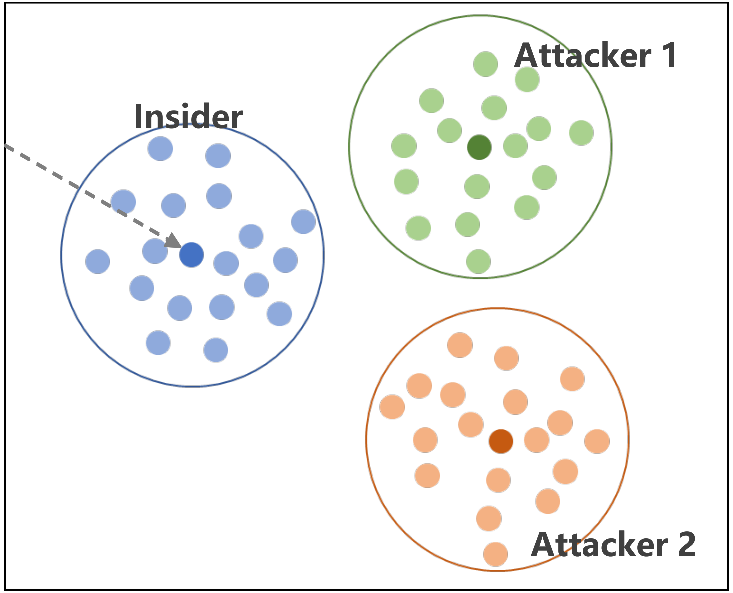


Average Face



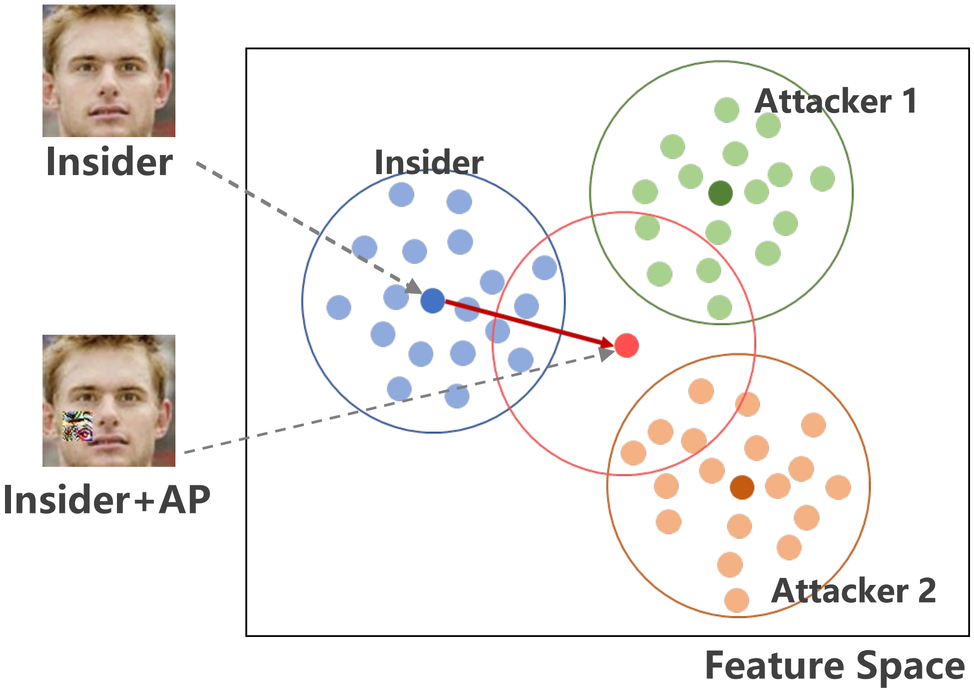
Insider

Insider



Feature Space

How to achieve UniID...



UniID still has challenges...

UniID still has challenges...

C1: For a specific insider, selecting attackers is important

UniID still has challenges...

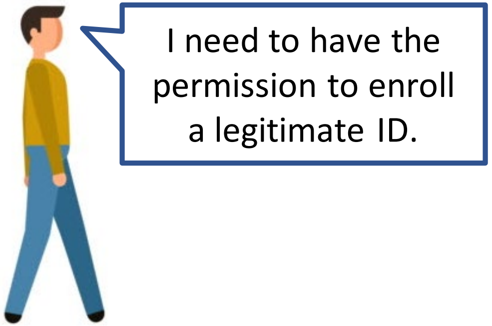
C1: For a specific insider, selecting attackers is important

I need to have the permission to enroll a legitimate ID.

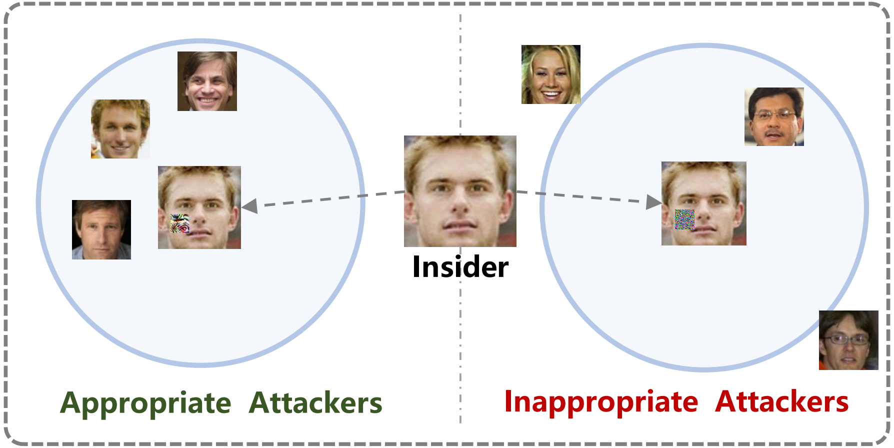
The choice of insider is restricted!

UniID still has challenges...

C1: For a specific insider, selecting attackers is important



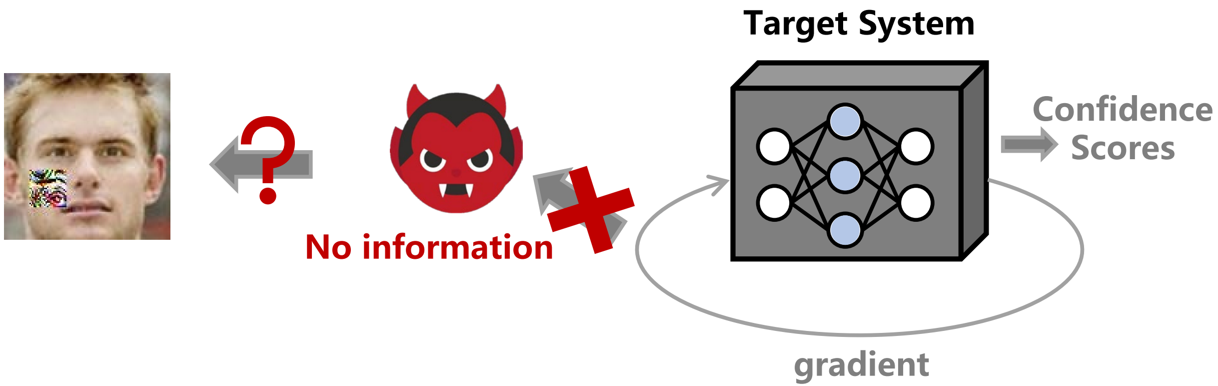
The choice of insider is restricted!



- ***Q1: How to determine the appropriate attackers?***
- *Q2: How to address the black-box setting?*
- *Q3: How to increase its physical robustness in real life?*

UniID still has challenges...

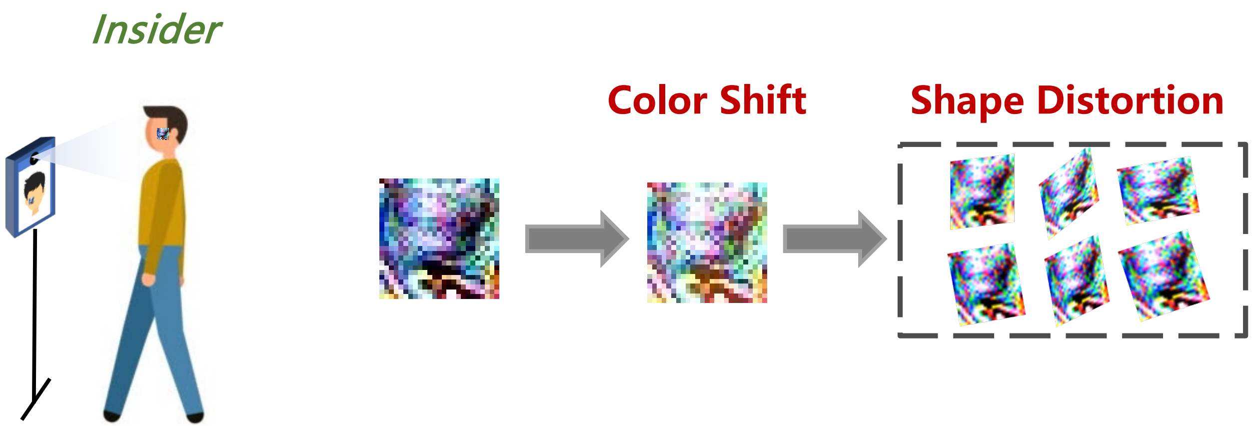
C2: Real-world face authentication systems are fully black-box settings



- ***Q1: How to determine the appropriate attackers?***
- ***Q2: How to optimize the adversarial patch under the black-box setting?***
- ***Q3: How to increase its physical robustness in real life?***

UniID still has challenges...

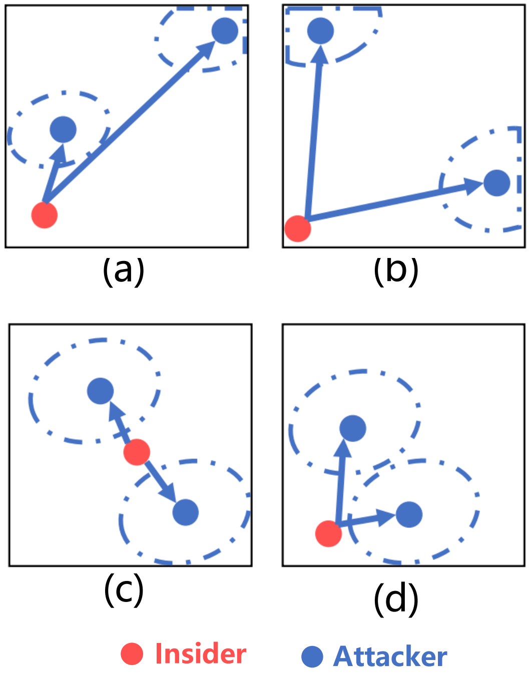
C3: The insider need to take photos on-site to upload his enrollment image



- ***Q1: How to determine the appropriate attackers?***
- ***Q2: How to optimize the adversarial patch under the black-box setting?***
- ***Q3: How to increase its physical robustness in real life?***

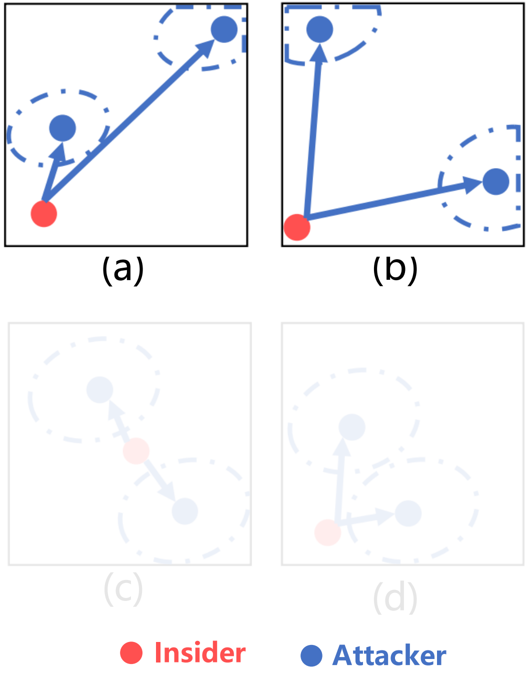
Q1 -> Attacker Selection

☐ *Multi-attackers Analysis:*



Q1 -> Attacker Selection

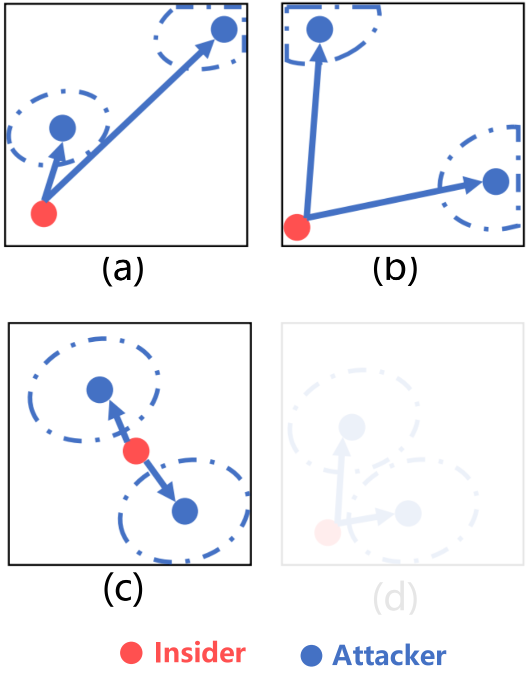
☐ *Multi-attackers Analysis:*



- **Case a & b:**
  The attackers are too far away from the insider

Q1 -> Attacker Selection

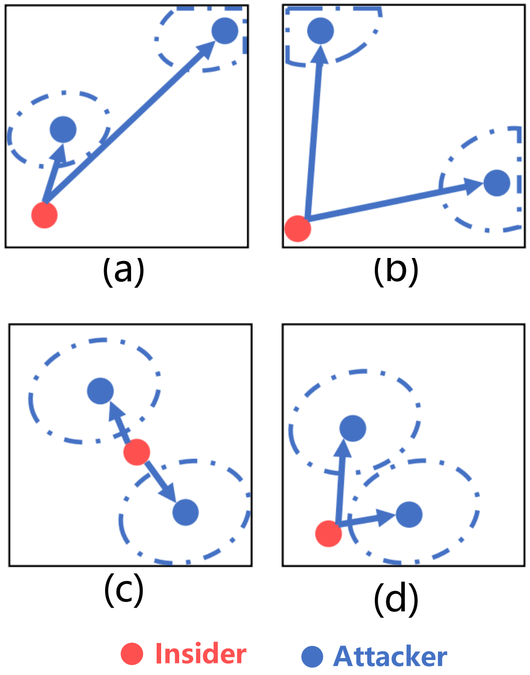
- ***Multi-attackers Analysis:***



- **Case a & b:**
  The attackers are too far away from the insider
- **Case c:**
  The attackers are located on either side of the insider

Q1 -> Attacker Selection

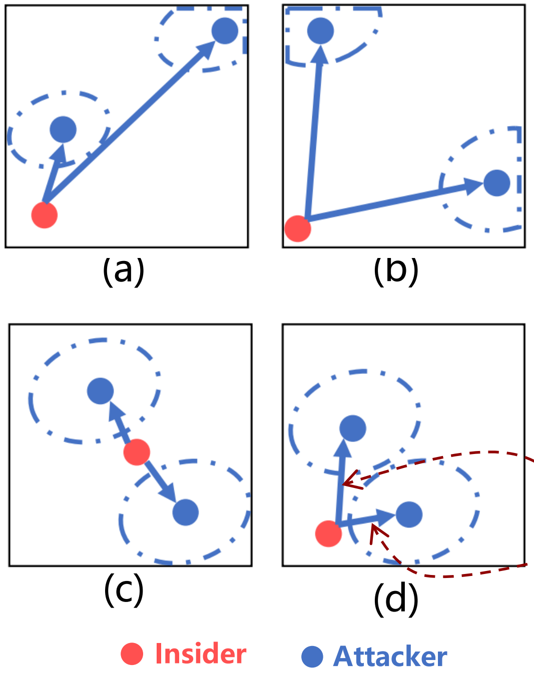
☐ ***Multi-attackers Analysis:***



- **Case a & b:**
  The attackers are too far away from the insider
- **Case c:**
  The attackers are located on either side of the insider
- **Case d:**
  The attackers and the insider are as close to each other as possible

Q1 -> Attacker Selection

☐ *Attacker Combination Choosing:*



- **Case a & b:**
  The attackers are too far away from the insider
- **Case c:**
  The attackers are located on either side of the insider
- **Case d:**
  The attackers and the insider are as close to each other as possible

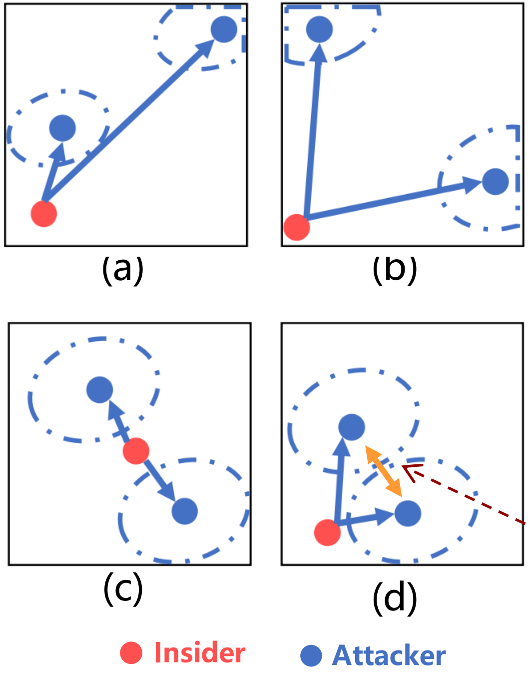
Similarity Metric

Aggregation Metric

$$Sim(V, A) = \frac{1}{N}\sum_{i=1}^{N} \frac{f(V) \cdot f(A_i)}{|f(V)| \times |f(A_i)|}$$

Q1 -> Attacker Selection

◻ *Attacker Combination Choosing:*



- **Case a & b:**
  The attackers are too far away from the insider
- **Case c:**
  The attackers are located on either side of the insider
- **Case d:**
  The attackers and the insider are as close to each other as possible

Similarity Metric

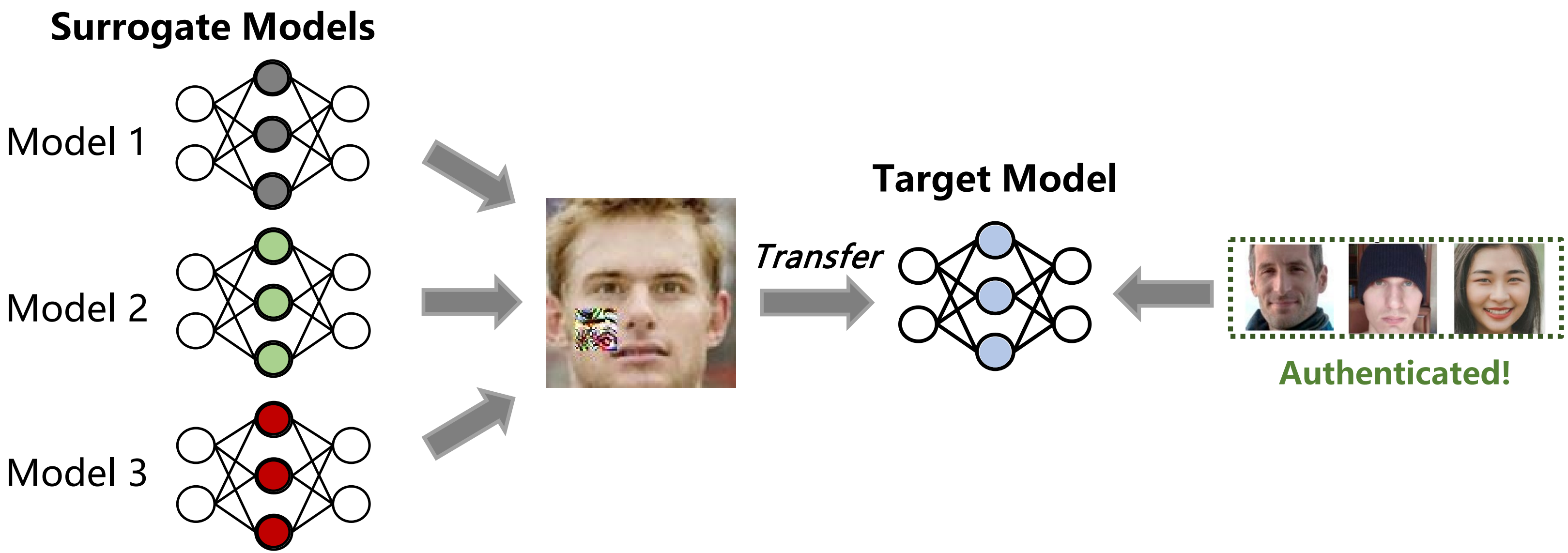
$$Sim(V, A) = \frac{1}{N}\sum_{i=1}^{N} \frac{f(V)\cdot f(A_i)}{|f(V)| \times |f(A_i)|}$$

Aggregation Metric

$$Agg(V, A) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{f(A_j)\cdot f(A_i)}{|f(A_j)| \times |f(A_i)|}$$

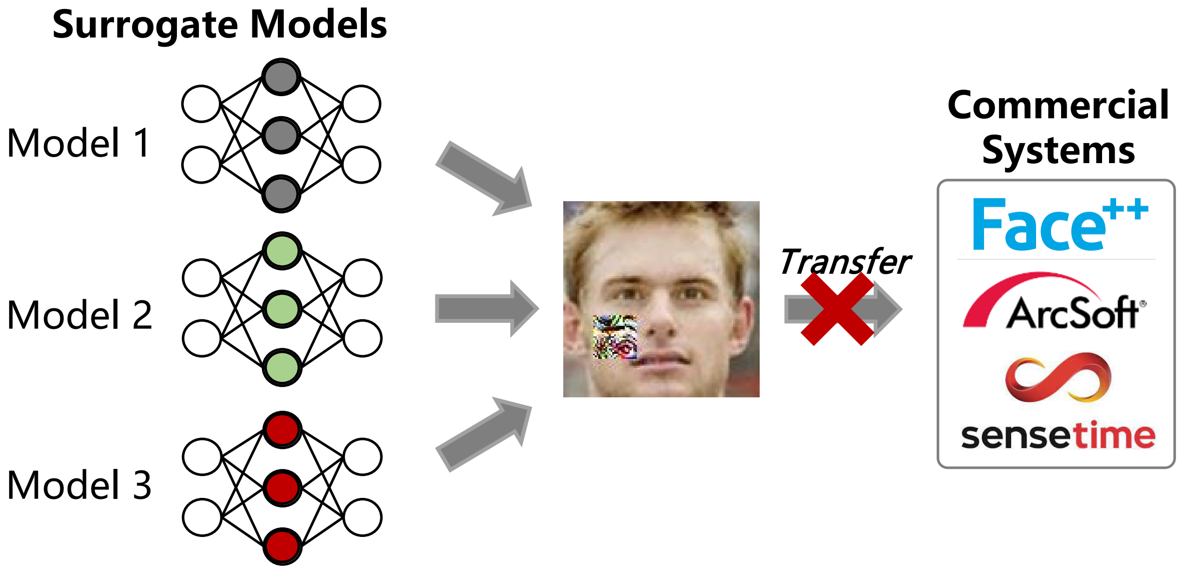
Q2 -> Black-box Transfer

- ***A straightforward method: Assembled-models***



Q2 -> Black-box Transfer

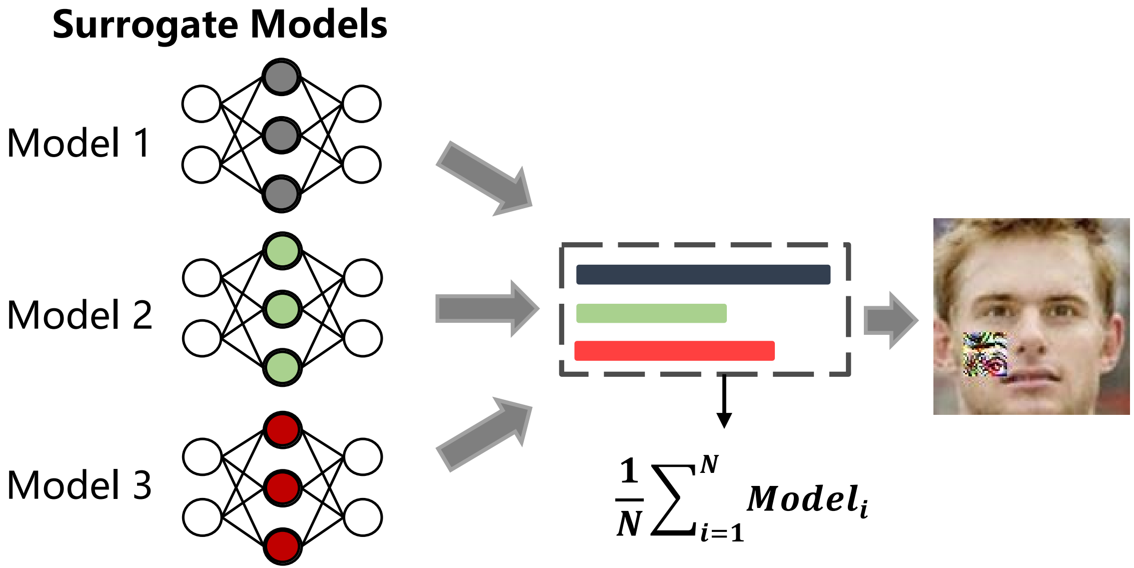
- ***The transferability is insufficient when targeting commercial systems***



- **High threshold**
- **Irregular models**

Q2 -> Black-box Transfer

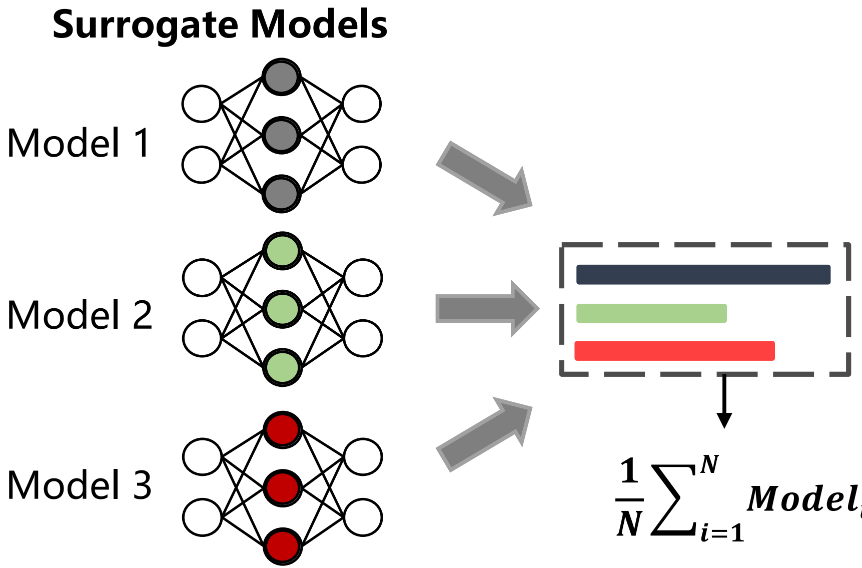
- ***Reason: Imbalanced gradients***



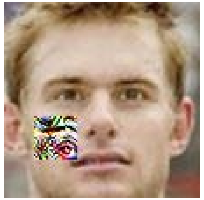
$$\frac{1}{N}\sum_{i=1}^{N} Model_i$$

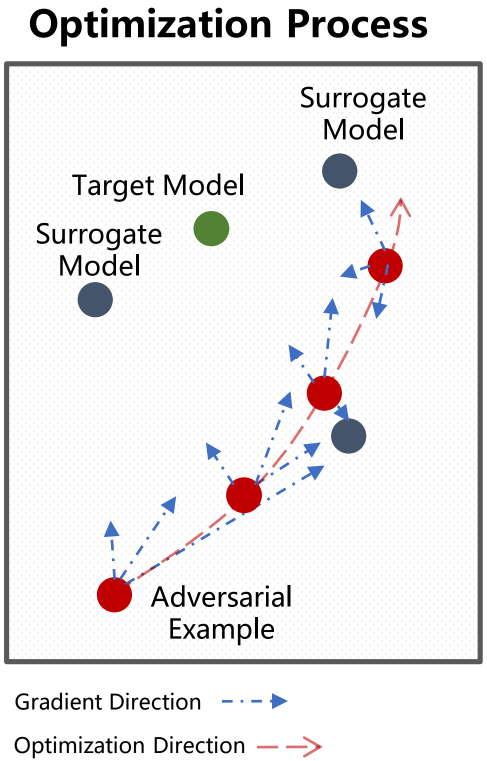
Q2 -> Black-box Transfer

- ***Gradient imbalance reduces effectiveness***



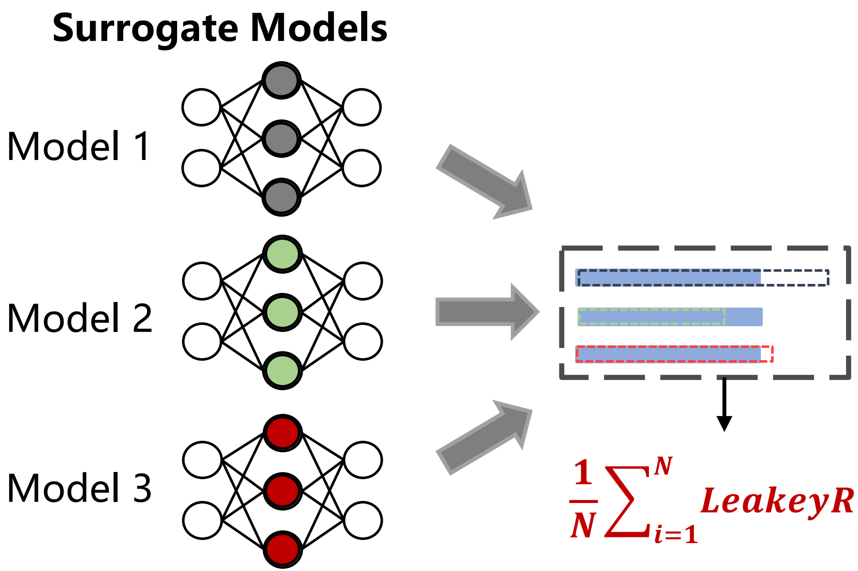
$$\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{Model_i}$$



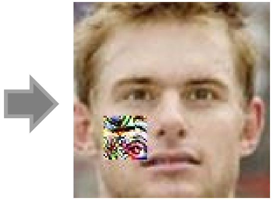


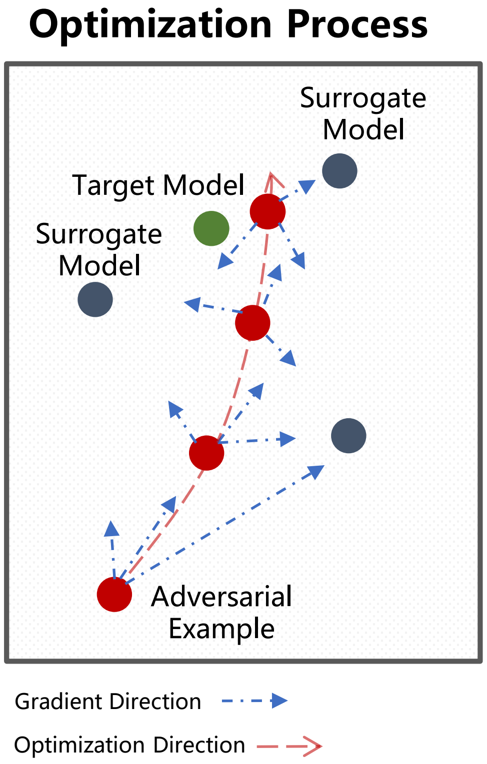
Q2 -> Black-box Transfer

□ *Agent Model Balance*



$$\frac{1}{N}\sum_{i=1}^{N} LeakeyReLU(Model_i)$$



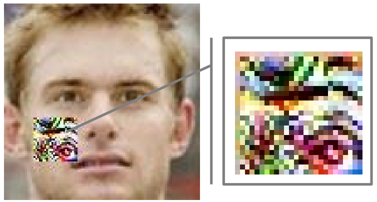


Q3 -> Physical Implementation

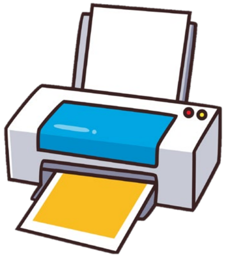
- *The printing-capturing process is a non-linear function*

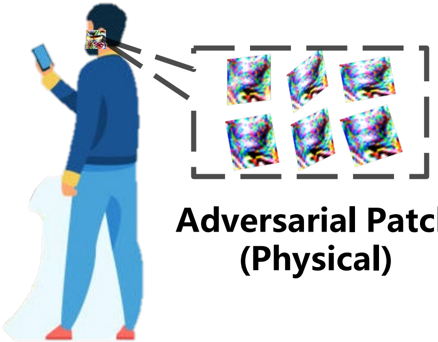
Printing Process

Capturing Process



Adversarial Patch
(Digital)





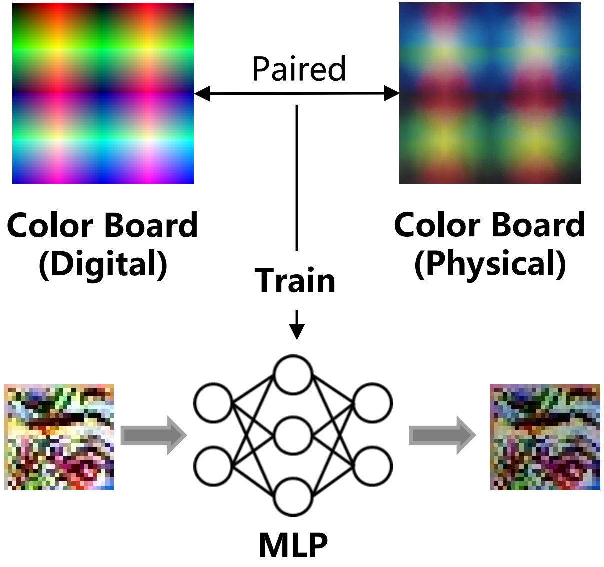
Adversarial Patch
(Physical)

Ink / paper / camera / face angle / light condition

Effectiveness

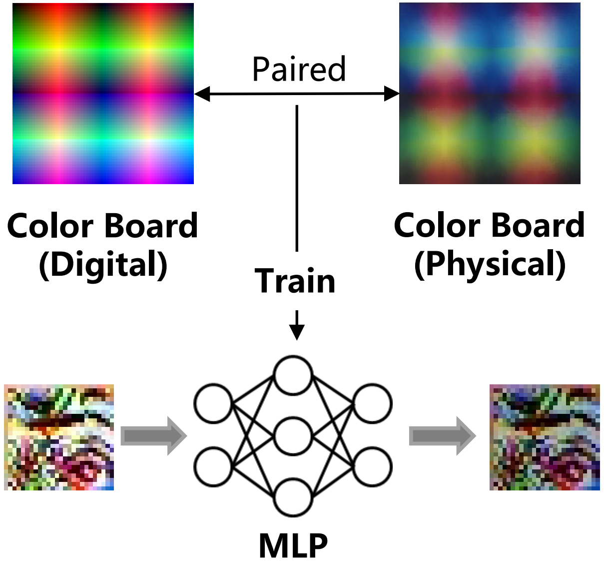
Q3 -> Physical Implementation

- ***Color-shift Calibration:***



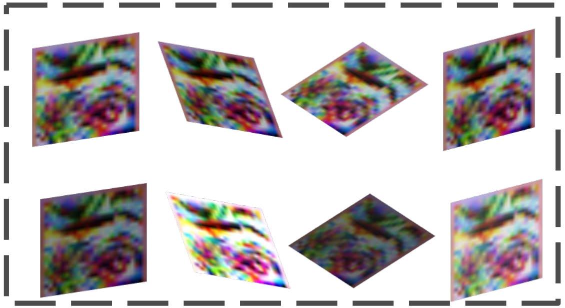
Q3 -> Physical Implementation

☐ *Color-shift Calibration:*



☐ *Shape-distortion Calibration:*

- **Expectation of Transformation (EoT)**



- Position
- Scaling
- Rotation
- Affine
- Brightness

Transform Distribution

$$\delta^* = \arg\min_{\delta} \mathbb{E}_{t\sim T}[\mathcal{L}[f(V, \mathbb{A}, t(\delta)), f(\mathbb{A})]]$$

The adversarial patch will be calibrated at each step of the optimization.

Evaluation

- ***Simulation Evaluation***
  - ***Overall Performance***
  - ***Impact of patch factors***
  - ***Impact of threshold settings***
- ***Real-world Evaluation***
  - ***Overall Performance***
  - ***Impact of light conditions***
  - ***Impact of camera settings***

Simulation Evaluation

☐ *Overall Performance :*

- ***Datasets: 100 users in LFW & CelebA***
- ***Target models: FaceNet, Mobile-FaceNet, ArcFace-18/50, MagFace-18/50, Face++, ArcSoft***

Table 1: Overall Performance in White-box Models

Target Models	Number of Attackers				
	1	2	3	…	7
FaceNet	99%	92%	81%	…	24%
M-FN	98%	80%	57%	…	8%
Arc-18	100%	99%	92%	…	46%
Arc-50	99%	83%	53%	…	1%
Mag-18	100%	99%	92%	…	53%
Mag-50	98%	81%	43%	…	2%

- ASR: The attack success rate

Under white-box setting

- **ASR: 100% in 3-Users Scenario**
  (1 Insider + 2 Attckers)
- **Can Extend to 8-Users Scenario**

Simulation Evaluation

❑ *Overall Performance :*

- ➢ ***Target models: FaceNet, Mobile-FaceNet, ArcFace-18/50, MagFace-18/50, Face++, ArcSoft***
- ➢ ***Datasets: 100 users in LFW & CelebA***

Table 2: Overall Performance in Black-box Models

Target Models	Number of Attackers		
	1	2	3
Arc-18	95%	79%	45%
Mag-18	98%	71%	36%
Mag-50	95%	62%	20%
Face++	81%	45%	20%
ArcSoft	86%	27%	12%

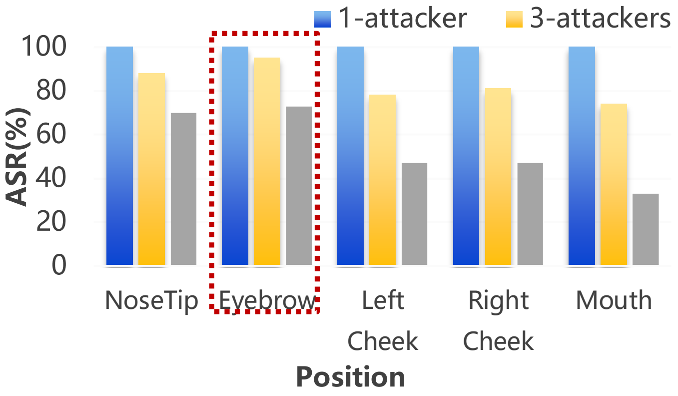
- ASR: The attack success rate

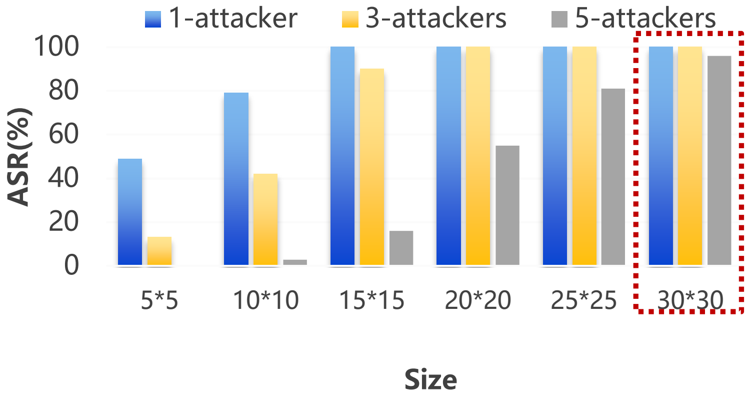
Under Black-box setting

- ➢ **ASR: 91% in 2-Users Scenario**
  (1 Insider + 1 Attckers)
- ➢ **ASR: 57% in 3-Users Scenario**
  (1 Insider + 2 Attckers)

Evaluation – simulation attack

- **_Attack Effectiveness:_**
  - **_Patch Position_**
  - **_Patch Size_**



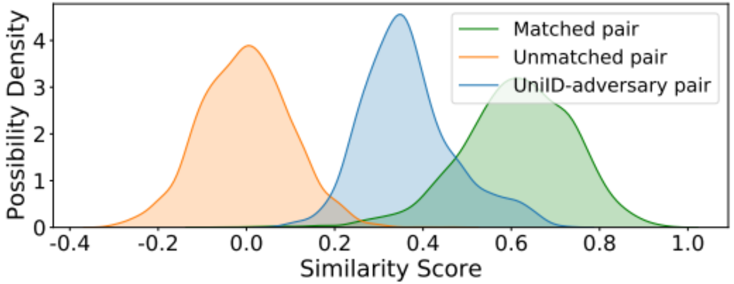


UniID is better to deploy in the eyebrow region with 30*30 size (7% of face)

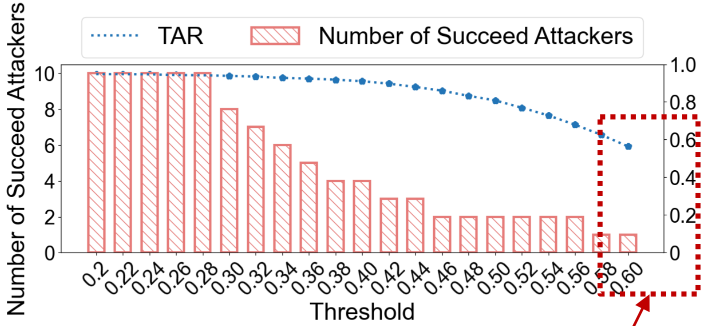
Evaluation – simulation attack

- **_Attack Effectiveness:_**
  - **_Threshold Setting_**

The distribution of similarity scores



Impact of different thresholds



40% of legitimate users are unable to authenticate

Merely increasing the threshold cannot simply block our attack

Real-world Evaluation

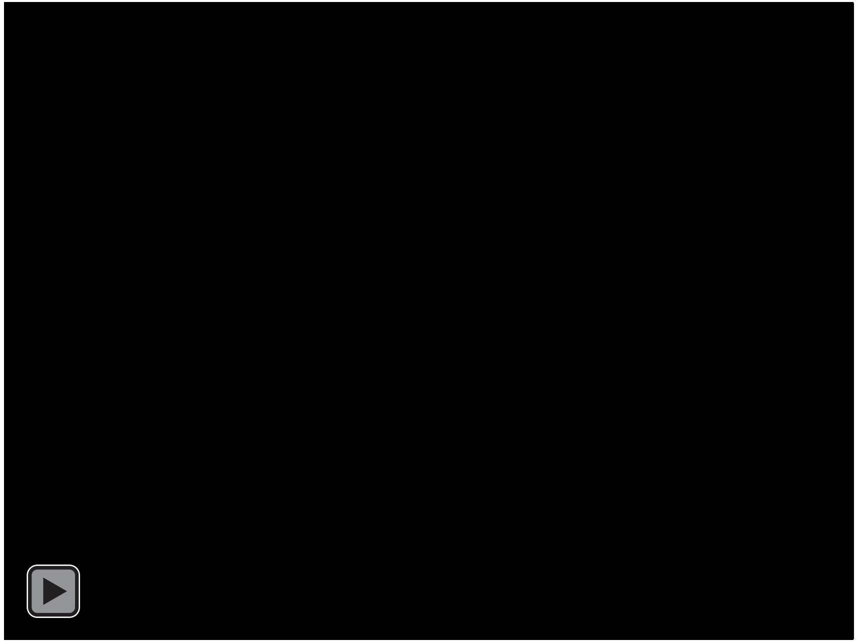
☐ *Overall Performance*

- ➢ ***Target system: Face++ & ArcSoft***
- ➢ ***Datasets: 20 volunteers***

Table 3: Overall Performance of UniID in Real World

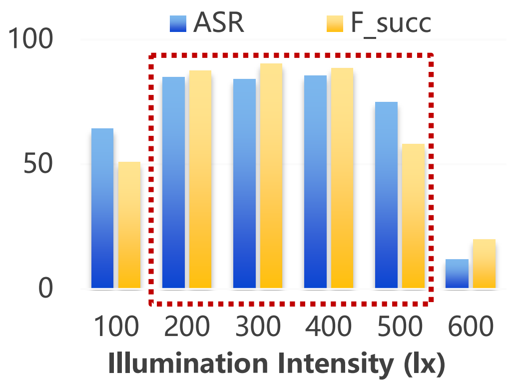
Metric	Target System	Number of Attackers	
		1	2
ASR	Face++	87%	41%
	ArcSoft	86%	47%
F_succ	Face++	84.3%	71.1%
	ArcSoft	86.5%	61.5%

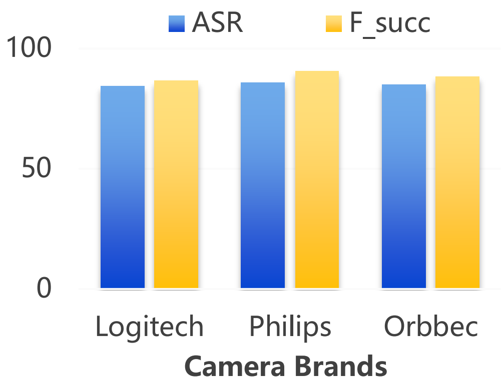
- ASR: The attack success rate
- F_succ: The attack success rate in consecutive frames

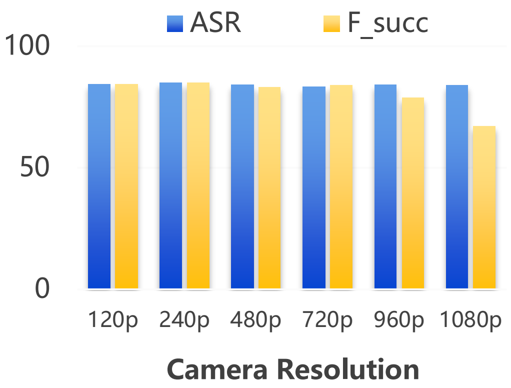


Real-world Evaluation

- ***Attack Effectiveness***
  - ***Light Conditions***
  - ***Cameras settings***







UniID is robust to various cameras in most light conditions

Discussion and Countermeasures

- ***Goal :***
  - *Offering **a systematic analysis** of face authentication security*
  - *urging service providers to **focus on security issues across all phases** of the workflow to make face authentication systems more secure*
- ***Countermeasures:***
  - *Enhancing the ability to distinguish different identities*
  - *Detecting adversarial examples at both the enrollment and recognition phases*
  - *Using assembled models to increase the attack difficulty*

Conclusion

- ***We identify the vulnerability in the face enrollment phase that enables multiple attackers to be successfully authenticated without any disguise.***

- ***We design UniID that make the legitimate user register a universal identity into the database, thus achieving the spoofing attack.***

- ***This vulnerability exists in other authentication systems that require an enrollment process.***

UniID: Spoofing Face Authentication by Universal Identity





Paper and demo website:
https://github.com/USSLab/UniID

Corresponding Authors:
yushicheng@zju.edu.cn
xji@zju.edu.cn

USSLAB Website: www.usslab.org