应对DeepFake实时图像流注入攻击风险

安全摄像头的探索与实践

张武　2024. 5

IIFAA

300家+
IIFAA联盟成员

16亿+
接入智能终端

900款+
覆盖手机型号

10亿+
服务用户数

关于IIFAA互联网可信认证联盟

成立于2015年的互联网可信身份认证联盟

IIFAA是蚂蚁集团在2015年联合中国信通院、阿里巴巴、华为、中兴、平安科技等联合发起的可信身份认证生态联盟。从2015年成立以来开拓了生物识别框架下的芯片级安全链路，普及了以人脸/指纹为代表的本地免密认证。目前，IIFAA联盟拥有300多家成员单位，覆盖了多元商业场景下领先的应用厂商、移动运营商、移动终端厂商、IoT厂商、芯片厂商、安全解决方案厂商、人工智能厂商、国家检测机构等“全产业链角色”。

全球超过16亿设备接入IIFAA技术平台

目前，IIFAA可信数字身份技术规范在全球超过**16亿台手机设备**、**43个手机品牌和900多款手机型号**上得到应用和支持；IIFAA还向支付宝、12306、建设银行、交通银行、东方航空、苏宁易购等金融、政务、线上购物和公共出行类应用提供服务，保障用户的安全体验。

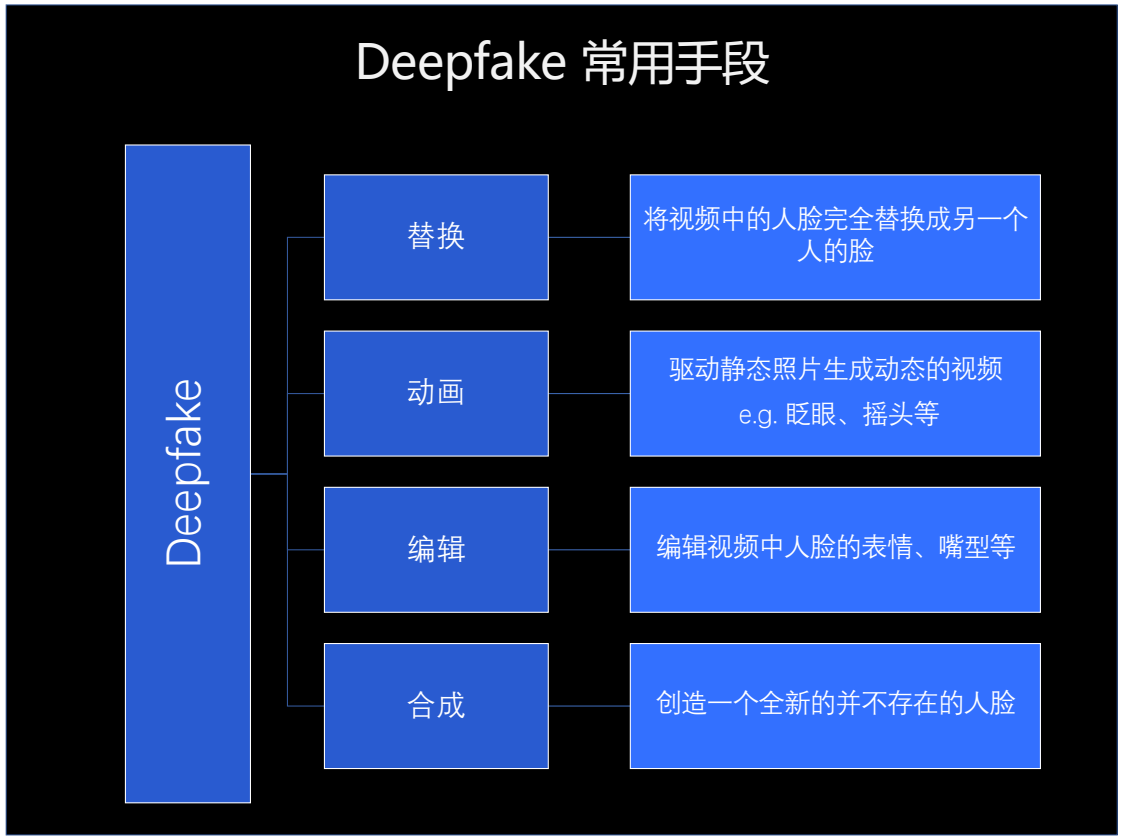
➢ DeepFake 技术演进及原理简介

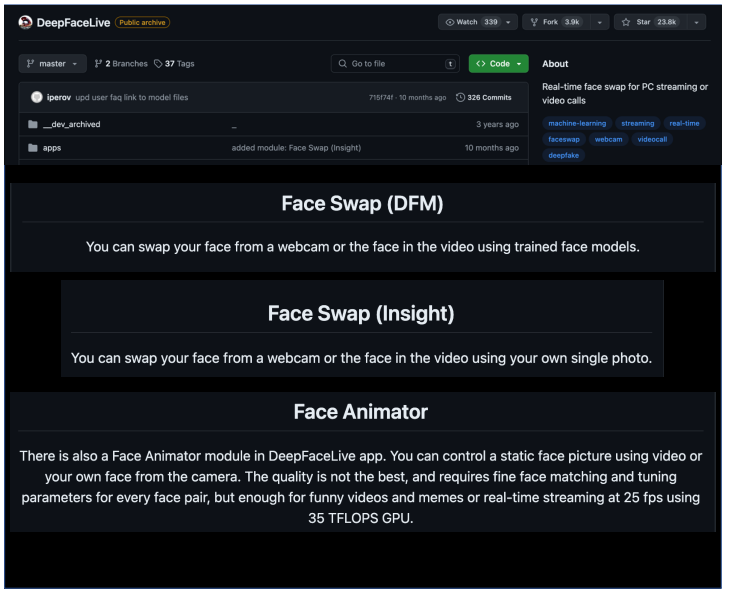
随机 AI 技术发展，深度伪造技术使用门槛越来越低，随之带来的隐私、欺诈等问题日趋严重，已经引起社会广泛关注

2014	2017	2018	2019	2020	2021	2022
Deepfake元年 (GAN 模型)	Reddit论坛出现首个人脸替换视频	FaceSwap 等工具普及，	Deepnude, ZAO 引起隐私和数据安全担忧	DeepFaceLab 开源	Diffusion models	Multi-model generation

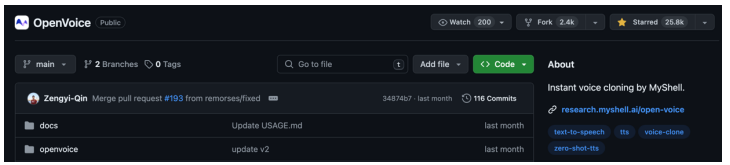
DeepFake 攻击原理及相关开源工具

结合开源工具，黑灰产可以极低成本的生成活体动作、视频对话等欺诈视频





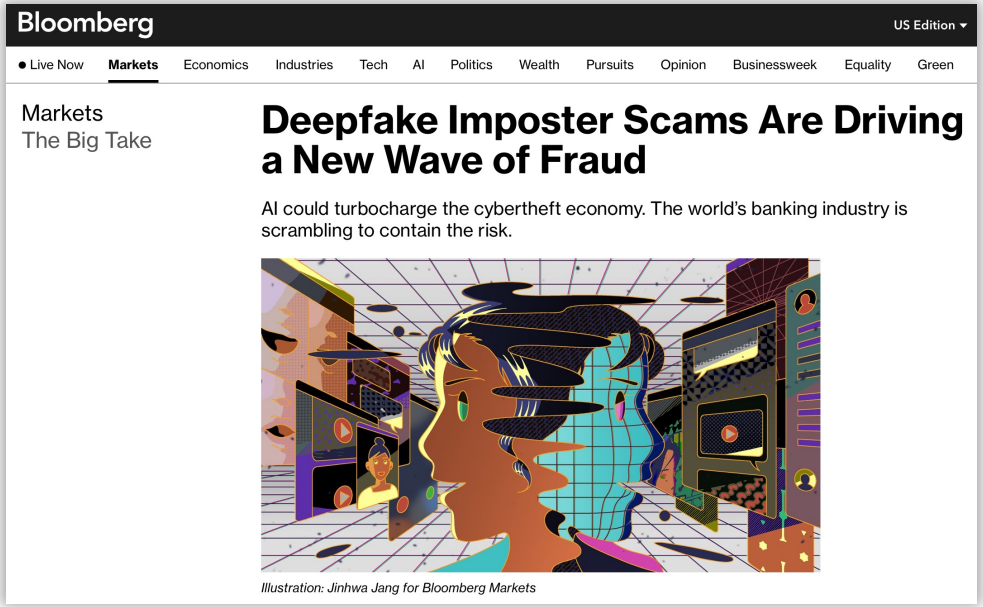
人脸DeepFake 开源工具：DeepFaceLive



声音Clone开源工具： OpenVoice

➢ 结合社工和 AI 技术的人脸欺诈攻击流程

攻击者通过社工等方式获取用户人脸、声音特征后，结合 AI技术生成虚假攻击视频，通过应用系统注入的方式进行欺诈攻击



Bloomberg

US Edition

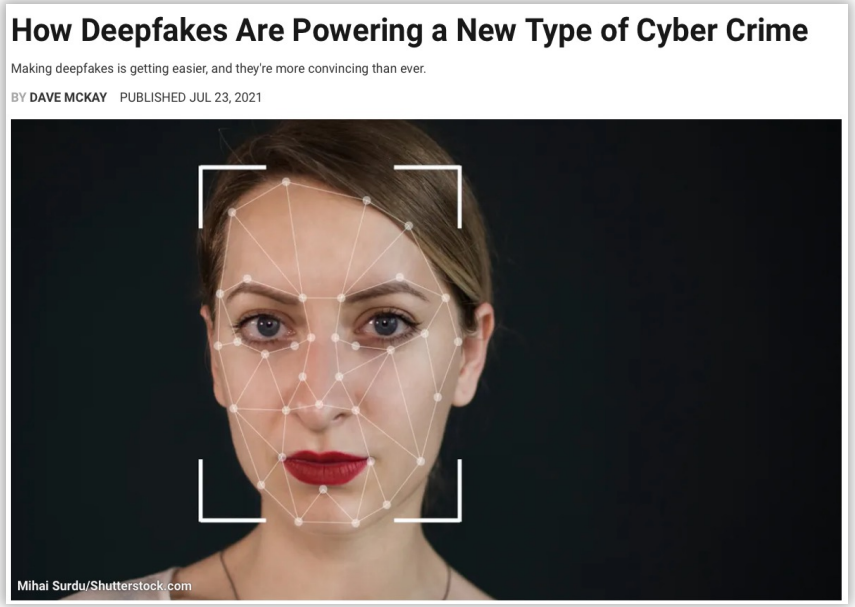
● Live Now | **Markets** | Economics | Industries | Tech | AI | Politics | Wealth | Pursuits | Opinion | Businessweek | Equality | Green

Markets
The Big Take

Deepfake Imposter Scams Are Driving a New Wave of Fraud

AI could turbocharge the cybertheft economy. The world's banking industry is scrambling to contain the risk.

Illustration: Jinhwa Jang for Bloomberg Markets



How Deepfakes Are Powering a New Type of Cyber Crime

Making deepfakes is getting easier, and they're more convincing than ever.

BY **DAVE MCKAY** PUBLISHED JUL 23, 2021

Mihai Surdu/Shutterstock.com

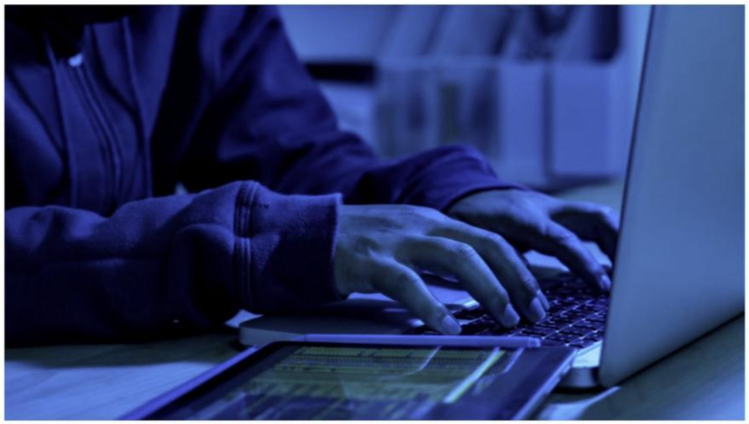
社交媒体 & 骚扰电话 → **获取用户人脸照片和声音** → **使用 AI 生成虚假视频** → **将虚假视频注入到业务流程进行欺诈攻击**

➢ DeepFake 实时视频流注入攻击案例：视频会议欺诈

一家跨国公司香港分公司的财务人员，按照“CFO”的指令，给对方指定账户共计转账2亿港币（约合1.8亿人民币）

Finance worker pays out $25 million after video call with deepfake ‘chief financial officer’

By Heather Chen and Kathleen Magramo, CNN
Published 2:31 AM EST, Sun February 4, 2024



boonchai wedmakawand/Moment RF/Getty Images

Authorities are increasingly concerned at the damaging potential posed by artificial intelligence technology.

(CNN) — A finance worker at a multinational firm was tricked into paying out $25 million to fraudsters using deepfake technology to pose as the company’s chief financial officer in a video conference call, according to Hong Kong police.

The elaborate scam saw the worker duped into attending a video call with what he thought were several other members of staff, but all of whom were in fact deepfake recreations, Hong Kong police said at a briefing on Friday.

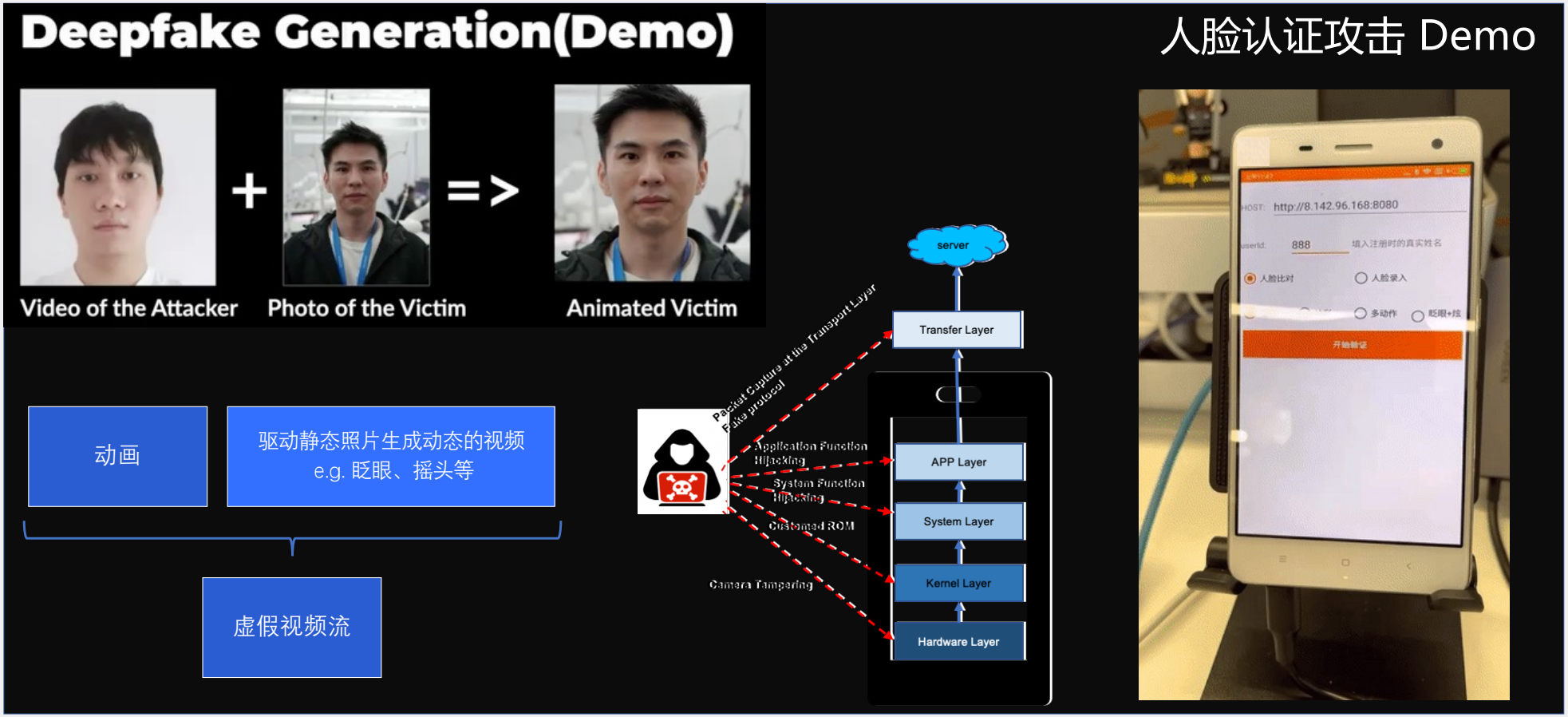
“(In the) multi-person video conference, it turns out that everyone [he saw] was fake,” senior superintendent Baron Chan Shun-ching told the city’s public broadcaster RTHK.

- 人脸替换：将视频会议中的人脸完全替换成“CFO”的脸
- 声音克隆：利用克隆后的音色模型实时生成“CFO”的语音

→ 虚假音视频流

➢ DeepFake 实时图像流注入攻击Demo：人脸认证攻击

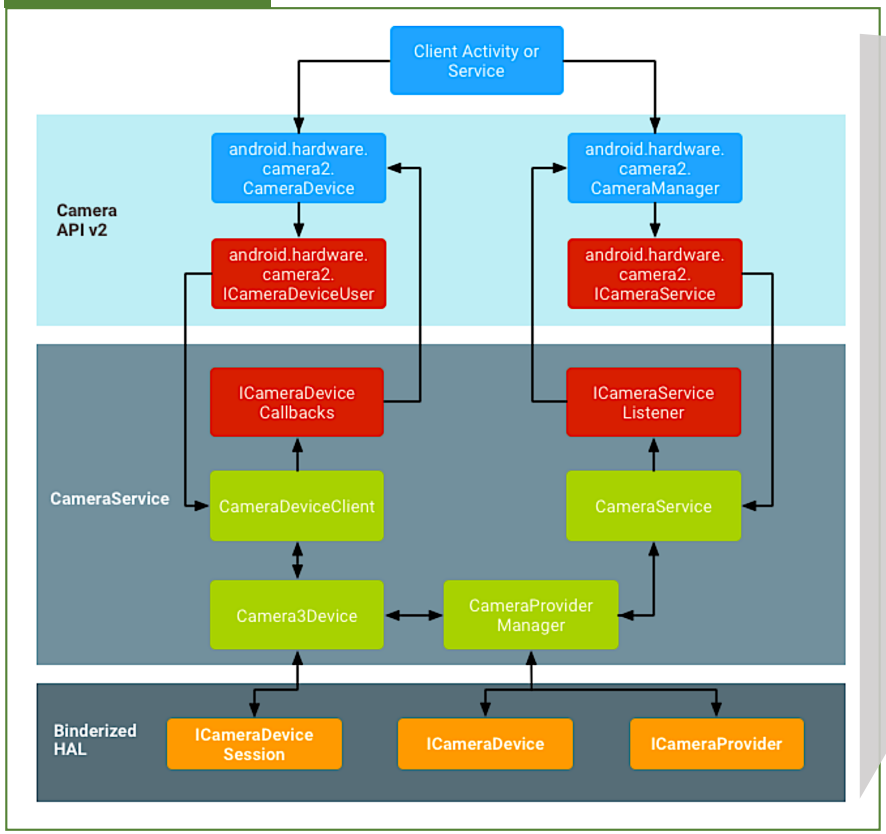
获取被攻击者人脸照片后，通过DeepFake 技术生成满足活体认证要求的人脸动作视频流，注入到身份认证流程进行账户盗用等攻击



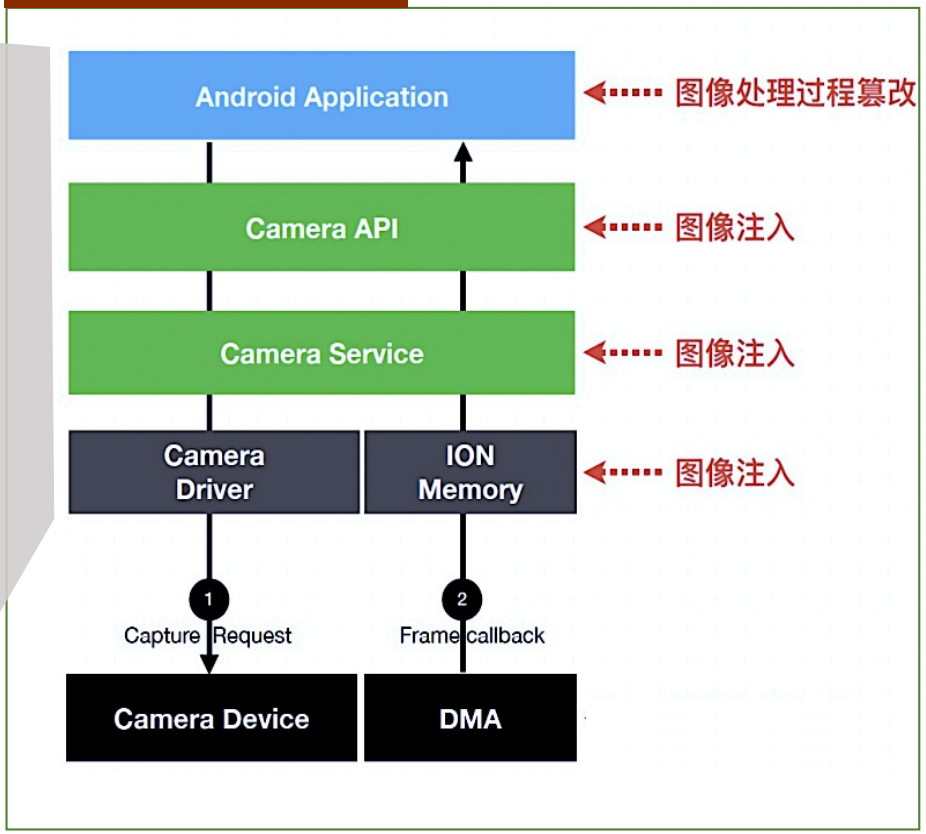
➢ 传统摄像头应用链路及风险分析

以 Android 系统为例，分析应用使用系统相机时的数据传输链路及潜在风险

Android 相机架构

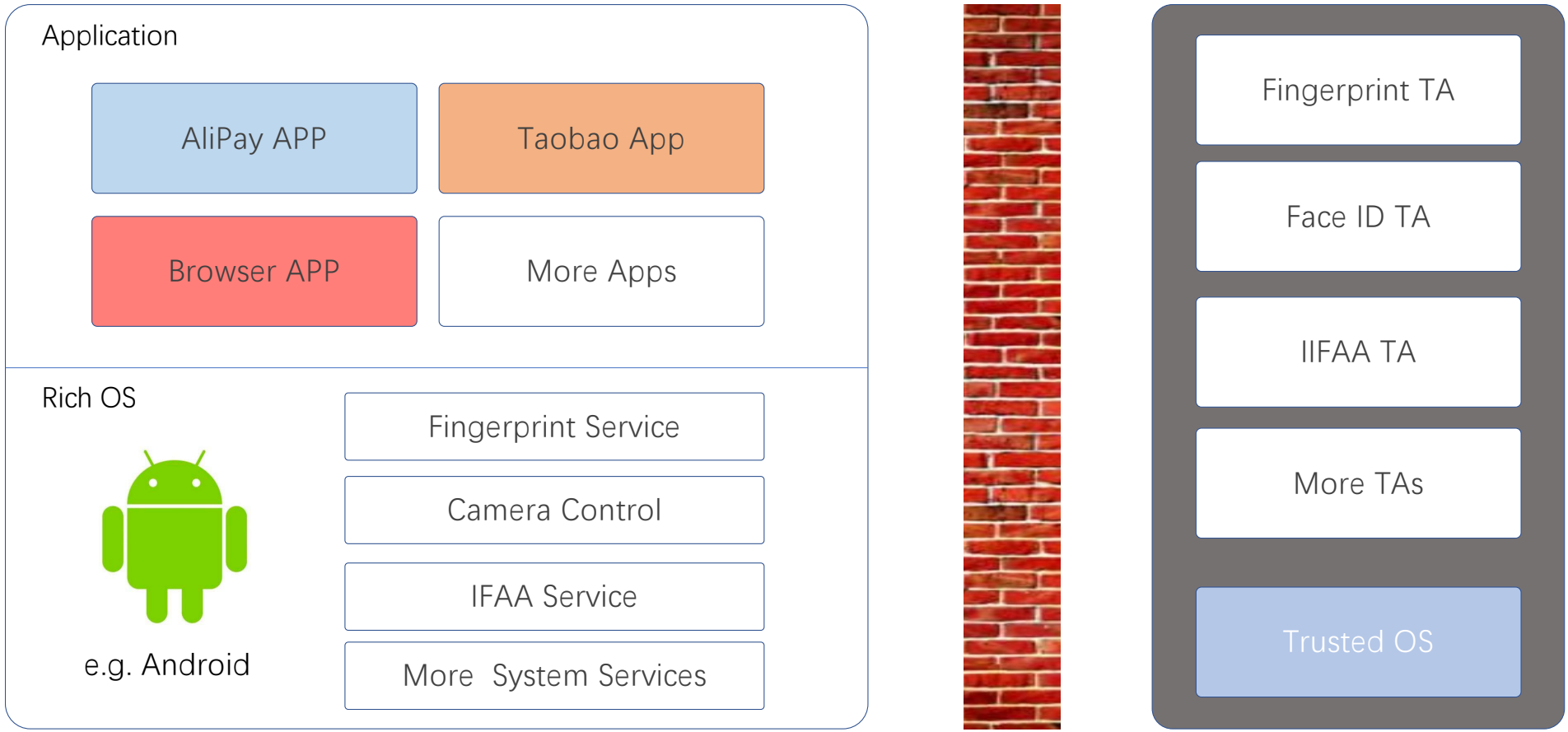


Android相机链路风险敞口



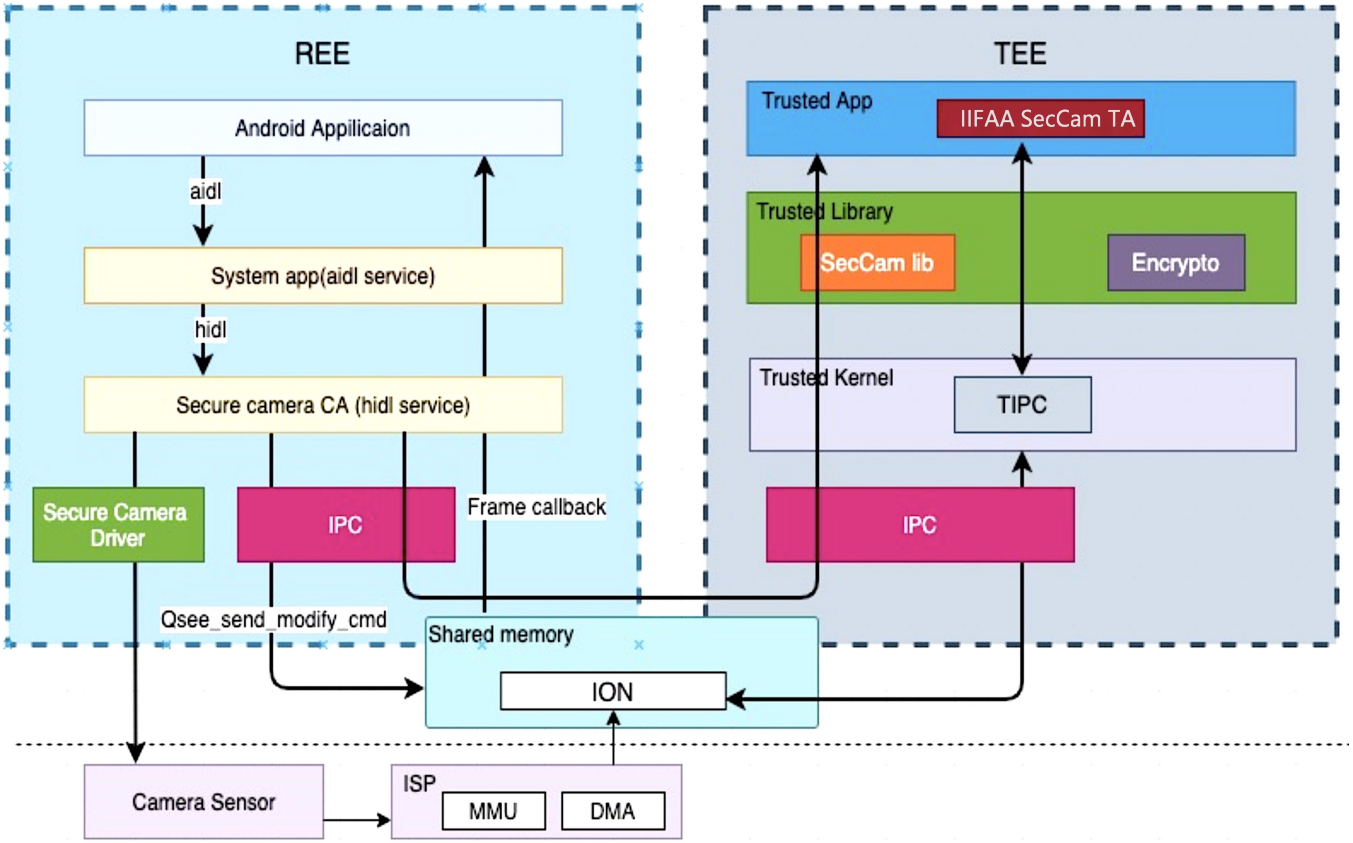
系统架构分析及安全方案探索

移动终端或 PC，普遍具备独立于富执行环境(Android/iOS等)的安全执行环境(TEE/TPM等)



➢ 安全摄像头技术方案探索

基于可信执行环境(TEE)的安全摄像头方案，通过在源头增加图片签名的方式，防止后续链路虚假/篡改图片注入攻击



1.安全摄像头启动时，底层开辟一块仅供TEE访问（REE侧无法访问）的安全内存

2.camera sensor 在通过IOMMU将图像数据传到安全内存

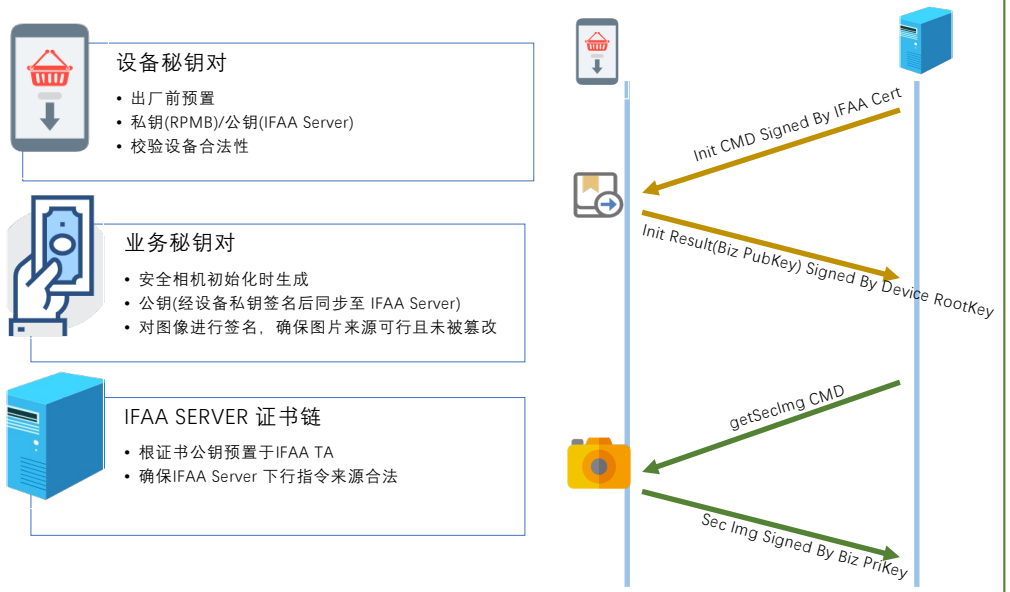
3.TEE中SecCam TA 通过能共享内存取得ISP 生成的图像数据

4.在TEE 中使用设备秘钥对图像数据加签名后回传给REE侧使用

安全摄像头安全协议和数据报文设计

基于可信执行环境(TEE)的安全摄像头方案，通过在源头增加图片签名的方式，防止后续链路虚假/篡改图片注入攻击

双向链路可信流程设计



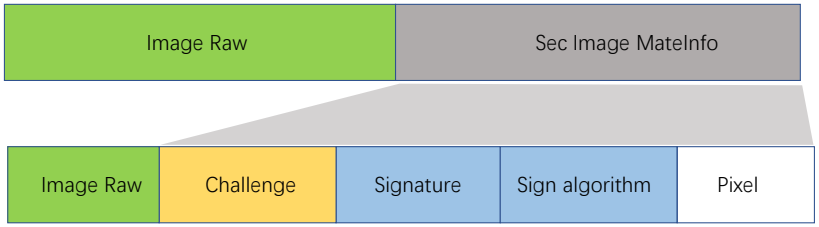
【可信端侧】与【认证服务器】各自预置私钥并持有对方公钥，通过签名机制确保双向链路数据交互可信

安全图像报文设计

1.初始化数据

2.初始化结果

3.安全图像数据

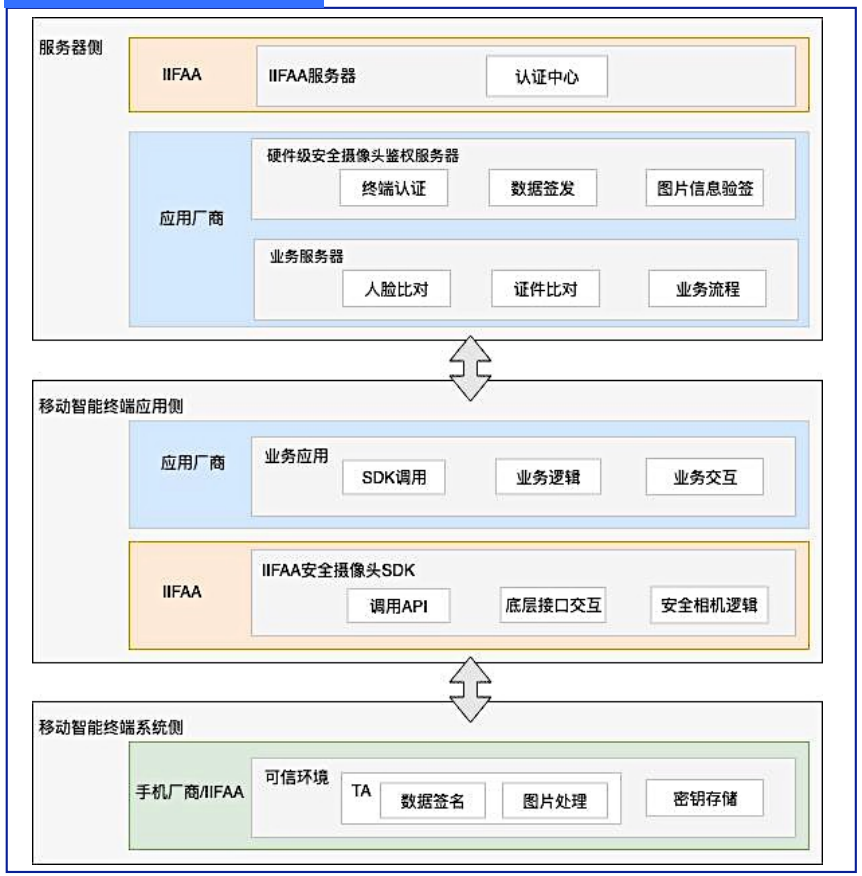


通过数字签名+挑战码机制，确保安全摄像头图像数据可信 且能有效防止重放攻击

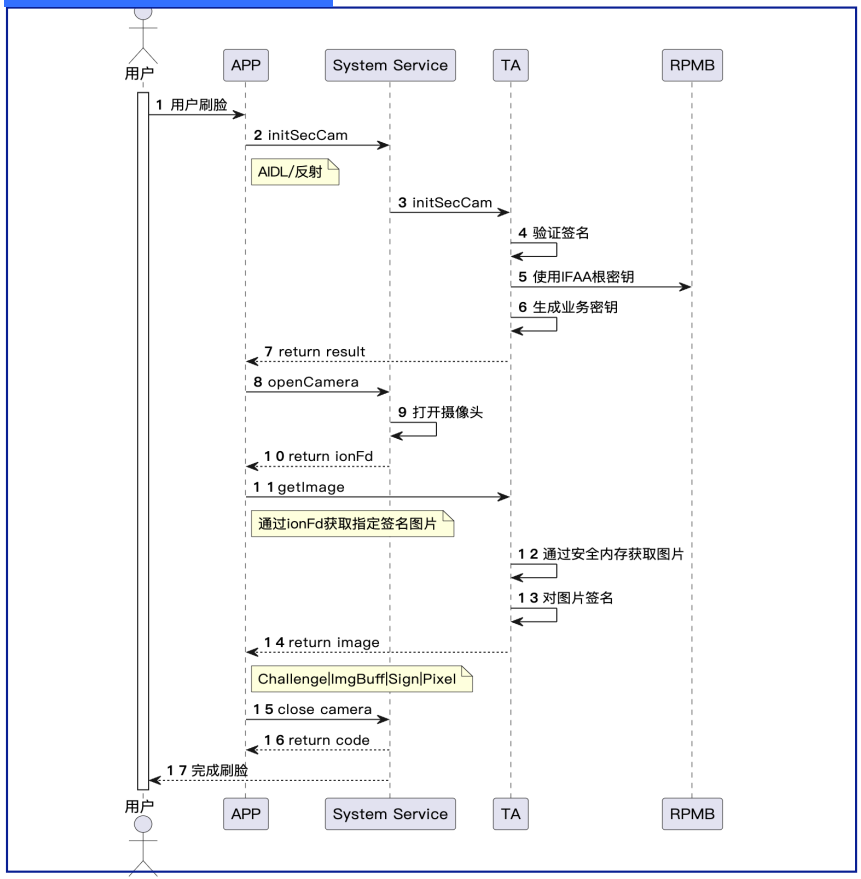
安全摄像头整体框架介绍及流程说明

基于移动终端进行了安全摄像头的整体设计及实现，通过使用 IIFAA 设备秘钥确保安全摄像头每一帧图像来源真实可信

安全摄像头架构设计



安全摄像头流程说明



IIFAA 标准成为多项国家和国际标准的基础

平台	标准	平台	标准
TC28/SC37国家标准工作组组长（2016~2022)	• GB/T370… • GB/T38…别安全，GB/T40668生物识别信息安全保护、GB/T41819…	…O / IECJTC1 / SC27 / WG5 建立…	• ISO/IEC27553-1:2022，ISO/IEC 27553-2、ISO 5158 3项 • ITU X.1451-2020 • IEEE 2790 -2020,IEEE P2859,IEEE P2884, IEEE P2891, IEEE P2866.1 5项

• YD/T 4064-2022 移动设备免密认证标准
• GA/T 1721\1722\1723 全系列 /1724/17…
CTID 行业标准

• …DAA 8 项团体标准；牵头数据安全团标
• TAF 001,002,011,012 移动设备指纹本地免密…标，096 SPTSM,110 AntDTX 等 6 项标准

护脸计划安全工作组副组长单位（申请中）

• 深度参与护脸计…业应用…

…识别检测、端侧可信解决方案落地和测评

IIFAA本地免密系列标准

1. T/IIFAA 0001本地免密2.0
2. T/IIFAA 0002 本地免密2.1 第1部分：架构
3. IIFAA本地免密 第2部分：IFAA安全域配置
4. IIFAA本地免密 第3部分：IFAA应用技术要求
5. IIFAA本地免密 第4部分：移动端管理接口技术要求
6. T/IIFAA 0003—2023IIFAA硬件级安全摄像头系统技术规范

3. 分布式认证技术规范第一部分：总体要求

3. T/IIFAA-3002.1-IIFAA远程声纹识别应用技术规范 第1部分：身份验证

IIFAA 硬件级安全摄像头系统技术规范

IIFAA硬件级安全摄像头系统技术规范

通过TEE环境采集安全图像并加签的技术，能在图像使用前确保其来自真实相机采集，极大提升了图像采集链路的安全性。明确了安全摄像头图片采集的流程及协议规范，进而提升了IIFAA本地免密技术规范的使用范围和安全水位。收起

T/IFAA

互 联 网 金 融 身 份 认 证 联 盟 标 准

T/IFAA 0003—2023

IIFAA 硬件级安全摄像头系统技术规范

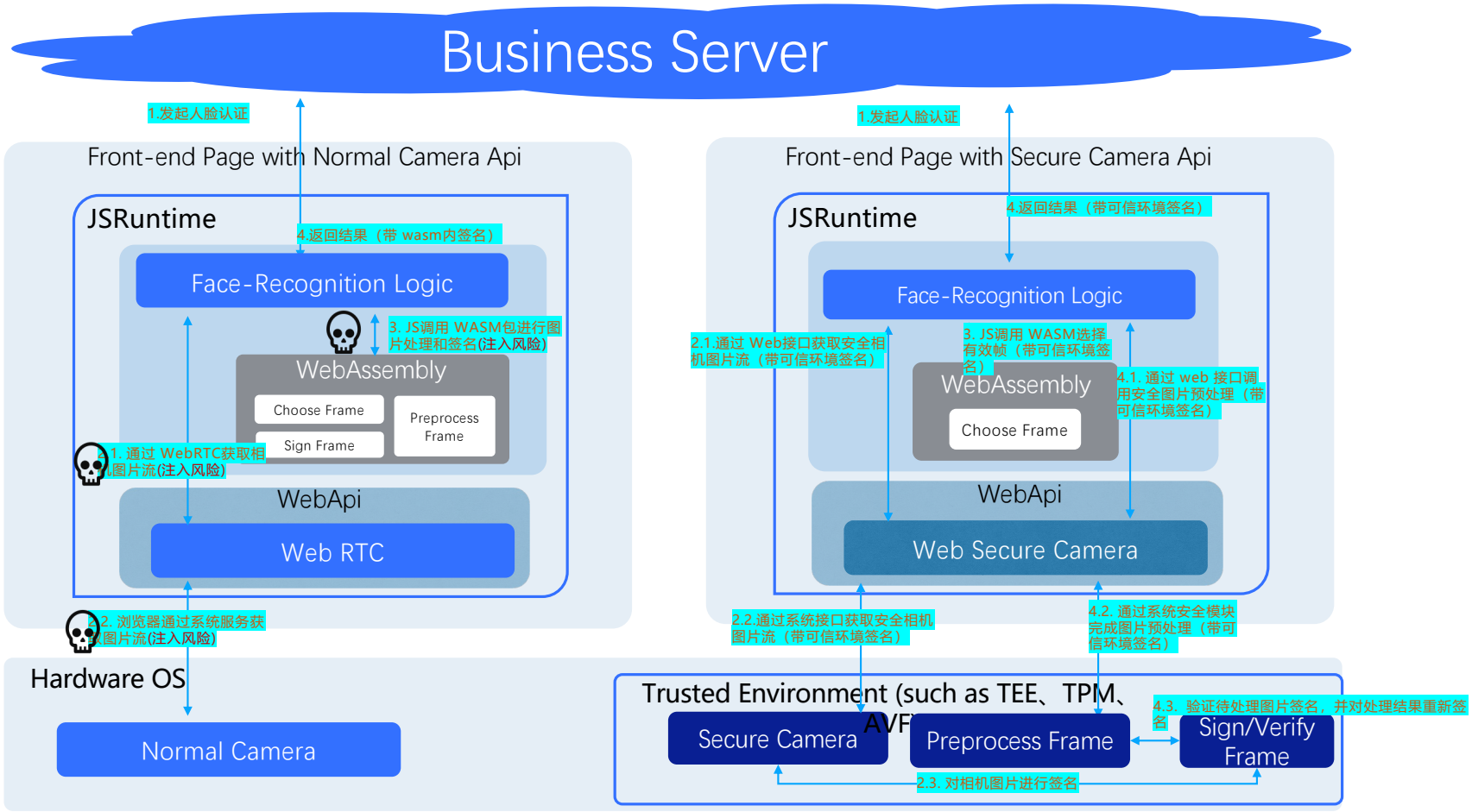
IIFAA Hardware-based Security Camera System Technical Specification

2023 - 01 - 04 发布 2023 - 01 - 04 实施

互联网金融身份认证联盟 发布

➢ 传统摄像头在 WEB 场景的风险分析及方案探索

以H5 刷脸场景为例，分析 WEB 链路图像注入攻击风险，结合 IIFAA 安全摄像头实践经验 探索WEB图像可信链路建设思路



共建 WEB 端安全摄像头标准

呼吁行业参与共建面向 WEB 生态的安全摄像头能力标准，构建真实可信的 WEB RTC 能力，提升生态安全水位

IIFAA可信摄像头标准已经在多家手机厂商采纳实现
并已面向行业移动应用开发者开放

IIFAA 联盟 呼吁 构建

面向 WEB 生态 的安全摄像头标准

构建统一标准的 WEB 端安全摄像头 API，降低开发者接入成本，提升行业安全水位

基于终端可信环境+数字签名等技术，生成源头可信的实时图像数据流

THANKS