Apple 燃起隐私计算

软件服务营收占比增加

- **Apple 营收可分为两部分：智能硬件+ 软件服务**
  - 软件服务是 Apple 第二大营收来源，仅次于 iPhone 硬件；
  - 较 2022 年，2023 年营收中仅软件服务实现正向增长 9%；

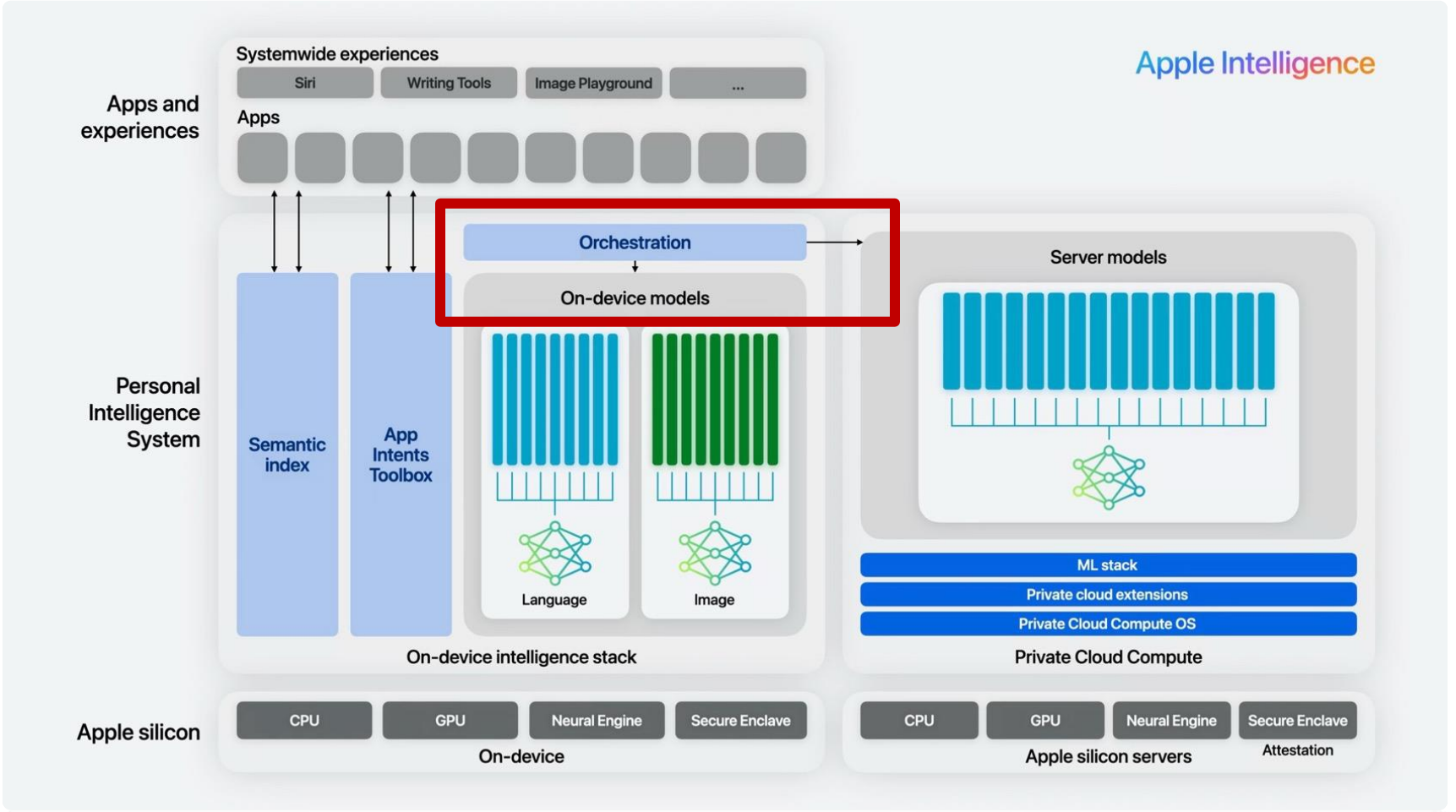
	Three Months Ended		Twelve Months Ended	
	September 30, 2023	September 24, 2022	September 30, 2023	September 24, 2022
(1) Net sales by category:				
iPhone	$ 43,805	$ 42,626	$ 200,583	$ 205,489
Mac	7,614	11,508	29,357	40,177
iPad	6,443	7,174	28,300	29,292
Wearables, Home and Accessories	9,322	9,650	39,845	41,241
Services	22,314	19,188	85,200	78,129
Total net sales	$ 89,498	$ 90,146	$ 383,285	$ 394,328

[Daring Fireball: Apple Q4 2023 Results]

1 背景：Apple 引燃

APPLE AI 端云协同

- 苹果计划采用一种结合设备端处理和服务器端处理的混合方式来实现人工智能功能。



APPLE AI 端云协同

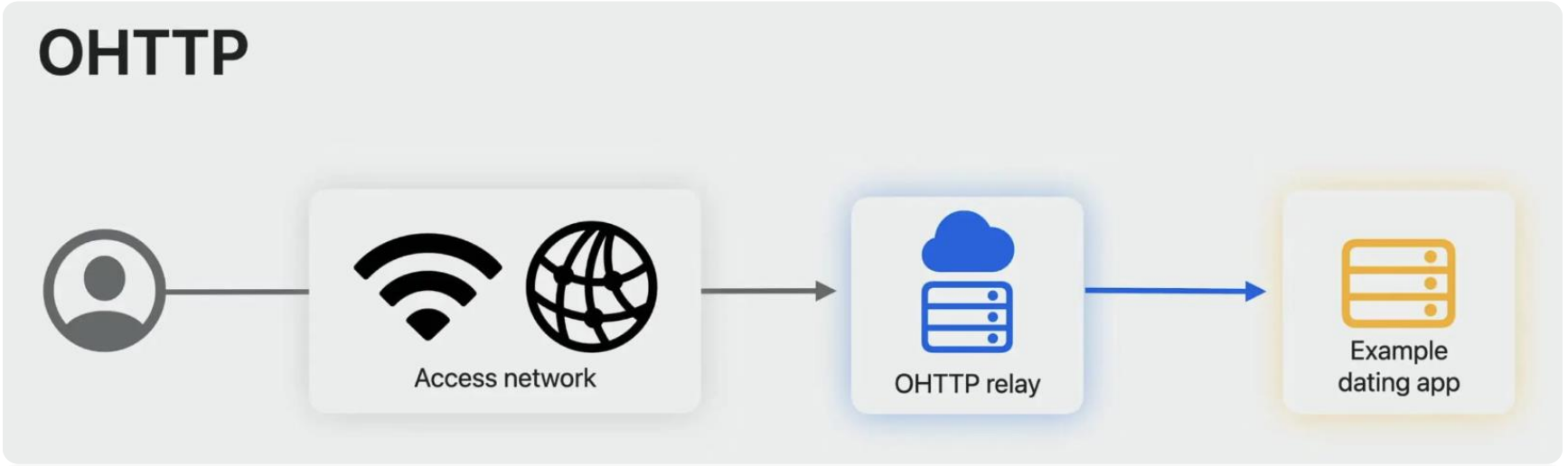
- Apple 与 OpenAI 合作，将 ChatGPT 能力内置于 iOS、iPadOs、MacOS，并面向 iPhone、iPad 和 Mac 的个人智能化系统 Apple Intelligence；
- Private Cloud Compute 通过 Private Cloud Compute，Apple 为 AI 的隐私功能树立了新标准，并能在设备端进程和搭载 Apple 芯片的更大型、基于服务器的模型之间灵活配置和扩充计算资源。

APPLE AI 端云协同

- 苹果宣称 PCC 是 AI 云端计算最先进的安全架构，目的比较明确，是要AI大模型时代，尤其是要端云协同的情况下，强化苹果安全隐私领导者的地位，让用户能放心使用整合在 i0S18 等 Apple Intelligence 功能。PCC 核心功能：
  - 用户数据以无状态方式进行计算，绝不存储用户数据；
  - 通过关键技术来保证隐私，不依赖外部服务；
  - PCC没有特权方式可以绕过隐私保护技术，也就是即使苹果 SRE 也没有办法接触用户数据；
  - 可验证安全隐私保护技术，可以开放给三方安全专家审核；

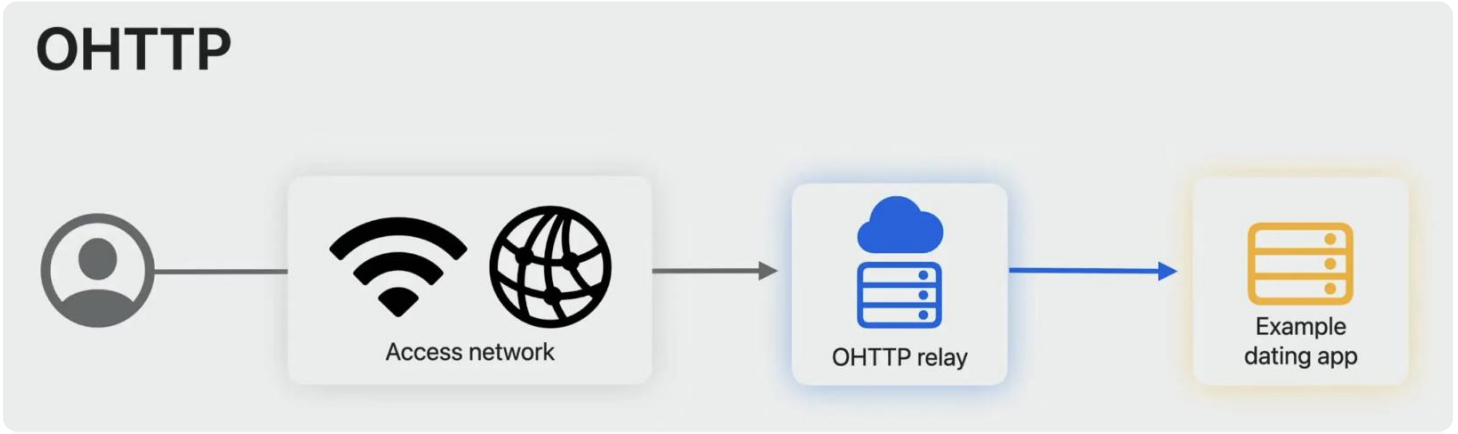
APPLE AI 端云协同：核心关键技术

- **安全云端环境：**云端经过苹果安全加固专用硬件和 OS，重点是苹果云端可信计算环境（类似TEE机密计算），苹果把 AI 重算力放到类 TEE 机密计算环境（类似软机密计算）；
- **消除通讯路径中风险点**：PCC 请求经过三方提供 OHTTP Relay 节点，OHTTP 方案在 WWDC2023 发布，可以隐藏源IP地址信息，引入 OHTTP 希望透明化运营商；



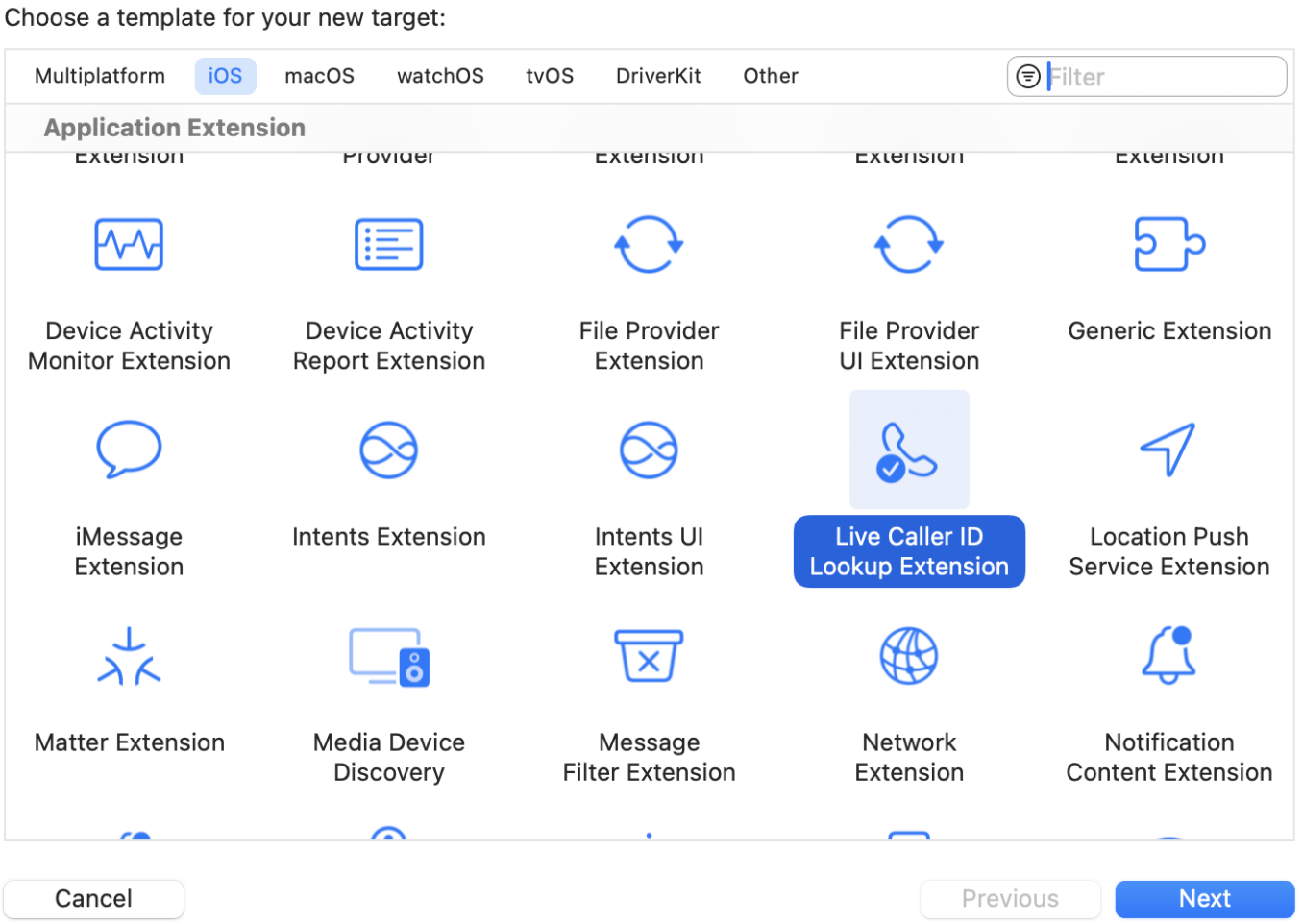
APPLE AI 端云协同：OHTTP

- 苹果 PCC 采用 OHTTP 进行隐私保护，Google 和其他浏览器厂商也在规划 OHTTP 方案，可以实现客户端、OHTTP代理、服务端，没有一个节点可以知道完整连接信息：
  - 客户端（知道源 IP 内容，不知道最终服务器地址）；
  - OHTTP 代理（知道源 IP、最终服务器地址，不知道内容）；
  - 服务器（知道 OHTTP 代理地址、内容，不知道源IP），安全隐私达到最大化。



APPLE 同态加密

- Apple 7 月 30 日宣布推出新的开源 Swift 包（swift-homomorphic-encryption），用于在 Swift 编程语言中启用同态加密技术。Apple 已经在最新 iOS18 中部署同态加密技术，典型应用实例 Live Caller ID Lookup.

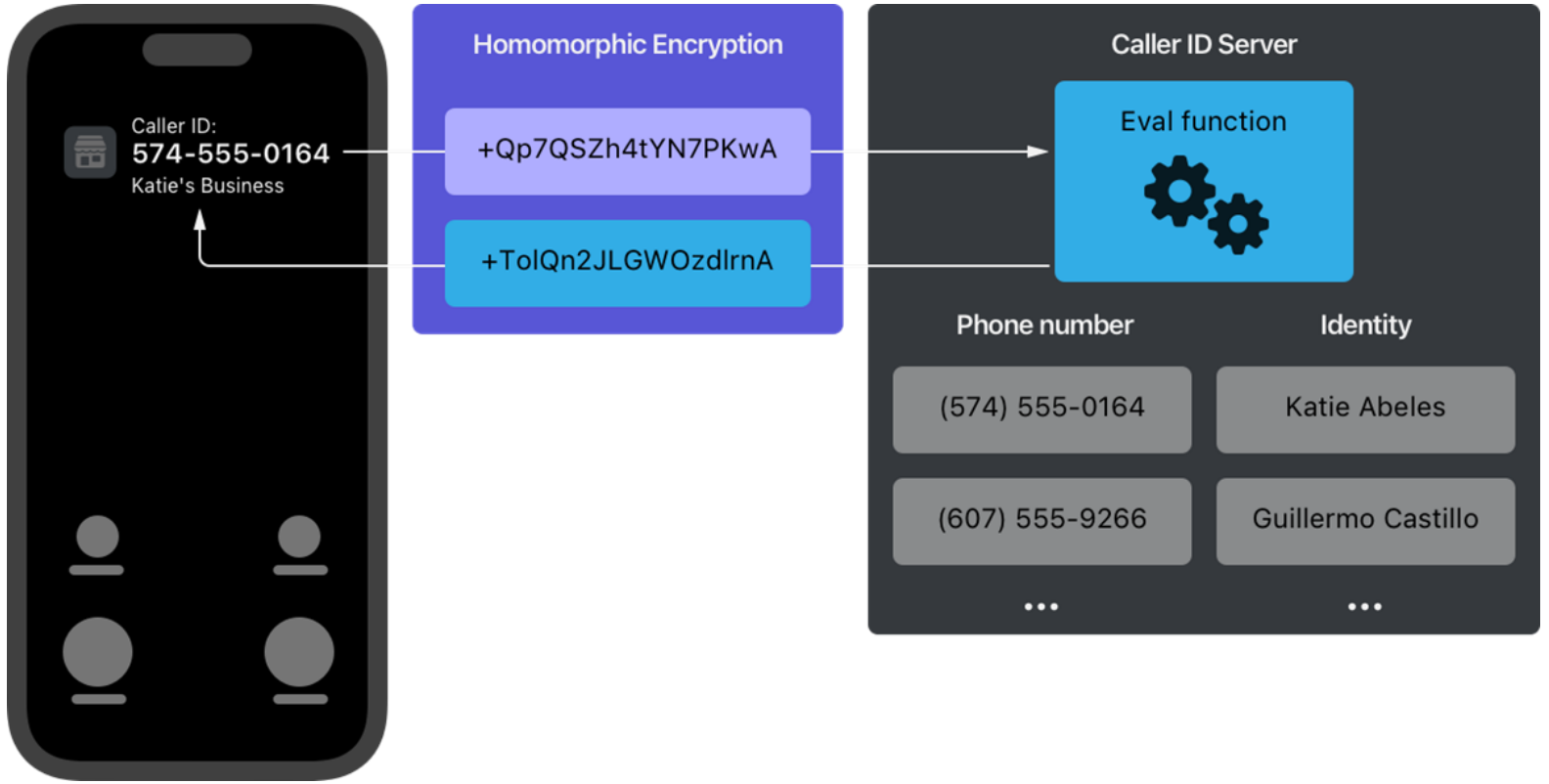


Live Caller ID Lookup：Swift-HE 实际应用

- **Live Caller ID Lookup（实时来电显示）**
  - iOS 18 的一个新特性，它使得系统与第三方服务进行通信可以有效保护用户信息，如隐藏 Client 的 IP 地址、进行匿名鉴别、隐藏来电号码等。
- **Private Information Retrieval（隐私查询）**
  - Live Caller ID Lookup 依赖于隐私查询（缩写为PIR）功能。PIR是一种私密键值对的数据查询形式，Client 可完成在 Server 数据库中查询而不泄露关键字（keyword）信息给 Server。

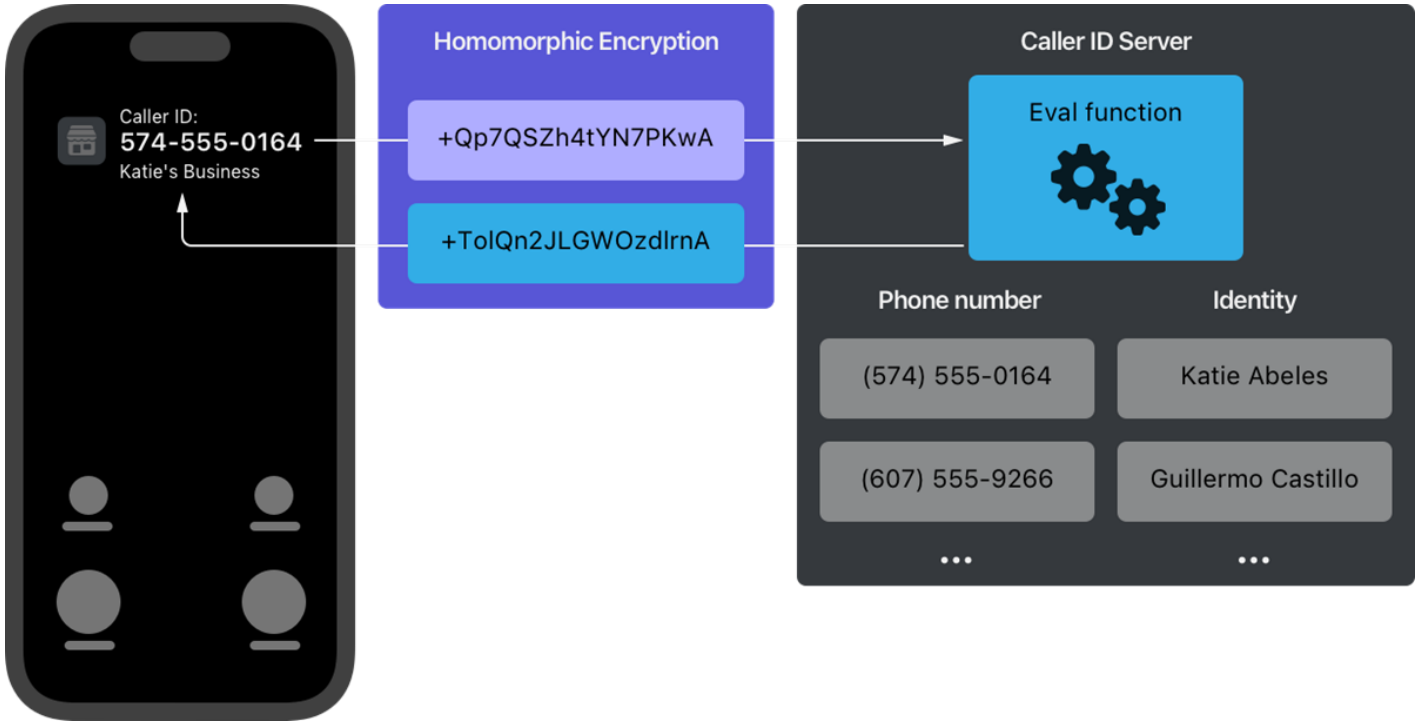
Live Caller ID Lookup：Swift-HE 实际应用

- Swift-HE 典型实例是拦截垃圾邮件和骚扰电话。当用户设备收到呼叫时，系统会与后端的服务器通信，以 PIR 方式检索呼叫者 ID 和拦截信息，同时在设备上显示对应的信息。具体包括：



Live Caller ID Lookup：Swift-HE 实际应用

- 使用Apple的不经意HTTP（Oblivious HTTP）隐藏Client的IP地址；
- 使用隐私通道协议（ Privacy Pass protocol ）进行匿名鉴别；
- 使用关键词隐私查询（ keyword PIR，KPIR ）隐藏来电号码。



Live Caller ID Lookup：Swift-HE 实际应用

- iOS18 使用 Swift-HE 实现这一功能，数据库和 Client 仅需同步小量元数据（metadata），无需频繁变化，使得处理流程相对高效。

https://github.com/apple/swift-homomorphic-encryption

2 技术：隐私计算

数据状态

- **数据状态：**静态（文件系统/数据库）、传输中（网络传输）、使用中（数据处理）。
- **数据生命周期：**数据存储、数据传输、数据计算、结果存储。
- **使用或处理数据时，通常会回到以下功能：**数据检索、数据分析、数据生成（包括 AI/ML）。

数据生命周期

- 根据数据生命周期，可将隐私计算参与方分为：输入方、计算方和结果使用方。
- 隐私计算应用中，至少有两个参与方，部分参与方可以同时扮演两个或两个以上的角色。

输入隐私 & 输出隐私

- 计算方进行隐私计算时需要注意“输入隐私”和“输出隐私”：
  - 输入隐私：参与方不能在非授权状态下获取 or 解析出原始输入数据以及中间计算结果；
  - 输出隐私：参与方不能从输出结果反推出敏感信息。

什么是隐私计算

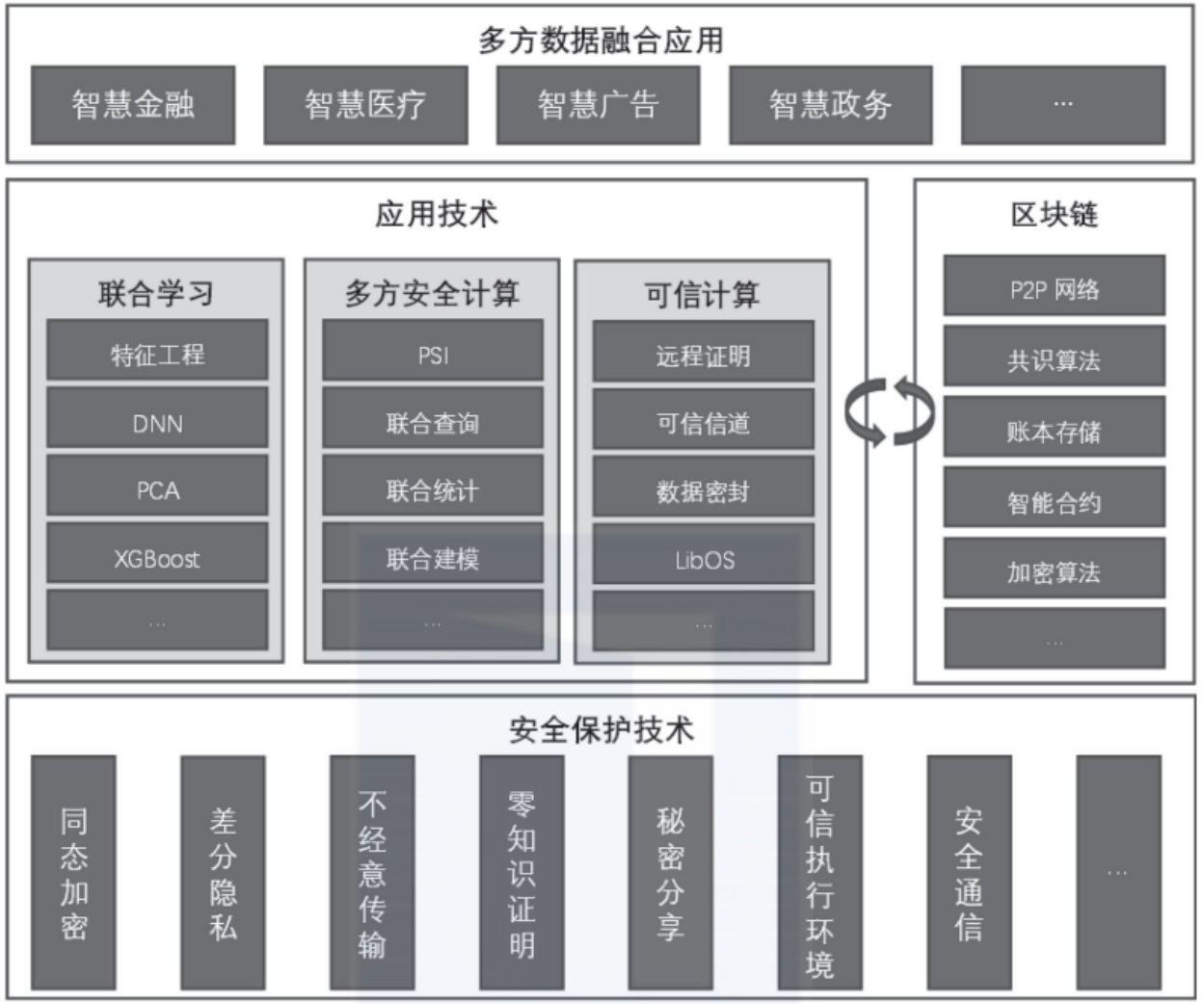
- **简单定义：**
  - 指在保证数据不对外泄露的前提下，由两个或多个参与方联合完成数据分析&计算相关技术的统称。
- **详细定义：**
  - 面向隐私信息全生命周期保护的计算理论和方法，具体是指在处理视频、音频、图像、图形、文字、数值、泛在网络行为信息流等信息时，对所涉及的隐私信息进行描述、度量、评价和融合等操作；形成一套符号化、公式化且具有量化评价标准的隐私计算理论、算法及应用技术，支持多系统融合的隐私信息保护。
- **作用：**
  - 隐私计算能够使上述功能以安全且私密的方式进行，从而允许以之前不可能的方式使用数据来释放价值。

隐私计算历史里程碑

[1] Shannon CE.Communication Theory of Secrecy Systems. 1949
[2] Rivest R, et al.On Data Banks and Privacy Homomorphisms. 1978
[3] Yao AC.Protocols for secure computations. 1982
[4] Goldwasser S, et al. The Knowledge Complexity of Interactive Proof-systems. 1985
[5] Paillier P. Public-Key Cryptosystem Based on Discrete Logarithm Residues. 1999
[6] Dwork C. Differential Privacy. 2006
[7] OMTP. Advanced Trusted Environment: OMTP TR1. 2009
[8] Gentry C. A Fully Homomorphic Encryption Scheme. 2009
[9] Wang S, et al. EXpectation Propagation LOgistic REgRession (EXPLORER). 2012
[10] Intel Architecture Instruction Set Extensions Programming Reference. 2015
[11] Konečný J, et al. Federated Learning: Strategies for Improving Communication Efficiency. 2016
[12] Gao D, et al. Privacy-preserving Heterogeneous Federated Transfer Learning. 2019

隐私计算全栈架构

- 虽然对于隐私计算技术的分类不同的看法，但几个核心支柱是确定的：
  1. 联邦学习
  2. 同态加密
  3. 安全多方计算
  4. 可信执行环境



隐私计算相关技术

	安全多方计算(MPC)	同态加密(HE)	差分隐私(DP)	零知识证明(ZK)	联邦学习(FL)	可信执行环境(TEE)
实现路径	• 数据系统与结构化	• 数据遮挡面	• 数据替换	• 数据系统与结构化	• 数据模型	• 数据隔离
保护数据阶段	• 使用状态数据(Data in Use)	• 存储状态数据(Data at Rest • 使用状态数据(Data in Use)	• 使用状态数据(Data in Use)	• 使用状态数据(Data in Use)	• 使用状态数据(Data in Use)	• 存储状态数据(Data at Rest • 使用状态数据(Data in Use)
优势特点	• 理论上无需第三方参与 • 直接得到结果和模型	• 无数据信息损耗	• 可根据需求添加“噪声”的量级，适用于多场景	• 达成特定目的且仅提供最低限度信息	• 原始数据不出库 • 分布式架构降低总算力成本	• 数据信息无损耗 • 域内无算法限制 • 现有成熟方案多
技术局限	• 所需算力较大 • 耗时长秘钥泄露存在可能	• 算力消耗大 • 随着数据增多，运算速度减缓显著，挤占网络带宽	• 添加“噪声”后，数据精准度下降	• 理论上安全性未被完全证实和广泛接受 • 某些类型的计算过程效 • 率较低标准化程度不高	• 数据模型质量参差不齐 • 通信复杂度较高 • 隐私保护无密码学验证	• 可能面临侧信道攻击
成熟度	• 产品化落地	• 已产品化、仍处于早期	• 产品化落地	• 产品化落地、方案不多	• 产品化落地	• 产品化落地、方案较多

隐私计算相关技术

- 同态加密瓶颈在于加密后数据传输量以及加密效率和性能。
- 零知识证明则成为后区块链时代重要武器。

隐私计算相关技术

关键技术	安全多方计算	联邦学习	可信执行环境
基本思想	基于密码学	数据不动模型动	基于可信硬件
性能	低到中	中	高
通用性	高	中	高
准确性	高	中到高	高
安全性	高	中	中到高
可信方	不需要	不需要	需要
性能	低到中	中	高
通用性	高	中	高
准确性	高	中到高	高
安全性	高	中	中到高
可信方	不需要	不需要	需要

隐私计算实现可信 AI

- 世界各地的组织都在寻找在不影响安全性的情况下实施 AI/ML 的方法，而这可以通过隐私保护机器学习 (隐私计算 + ML) 来实现。
- 隐私保护ML提供了一条创新途径，可以提取关键见解并推动AI/ML协作工作，同时保留 IP 和必要的数据敏感性要求和合规标准。
- 隐私计算通过两种重要方式为更广泛的ML领域做出贡献：通过推理和训练期间保护模型，允许组织将重点放在派生结果的业务收益上，而不是 ML 模型本身固有的风险上。

Thank you

把AI系统带入每个开发者、每个家庭、每个组织，构建万物互联的智能世界

Bring AI System to every person, home and organization for a fully connected, intelligent world.

ZOMI

Course chenzomi12.github.io

GitHub github.com/chenzomi12/AIFoundation