聚·变

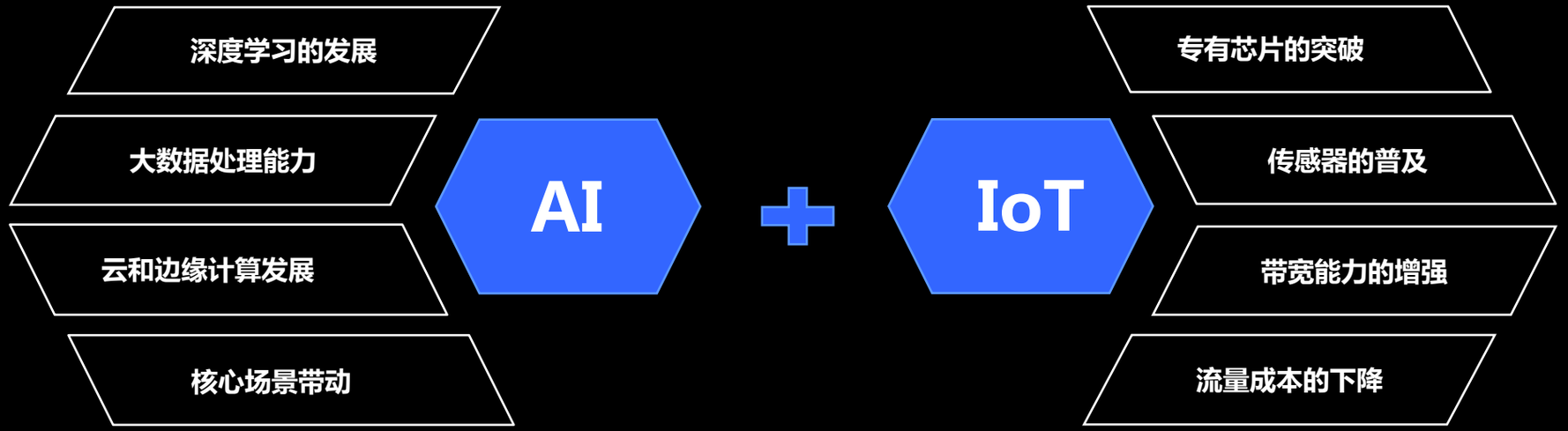
第二届顺丰信息安全峰会分论坛

AI 安全与隐私保护（1）

AIoT生态安全的新仇与旧恨

聂科峰　百度AI安全技术总监

AI加速IoT发展，进入全新的AIoT时代

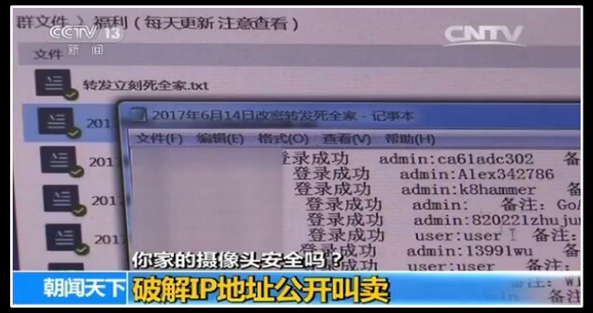


AIoT将物联网设备带入以感知、理解和自学习为特征的智能设备时代

AIoT安全问题

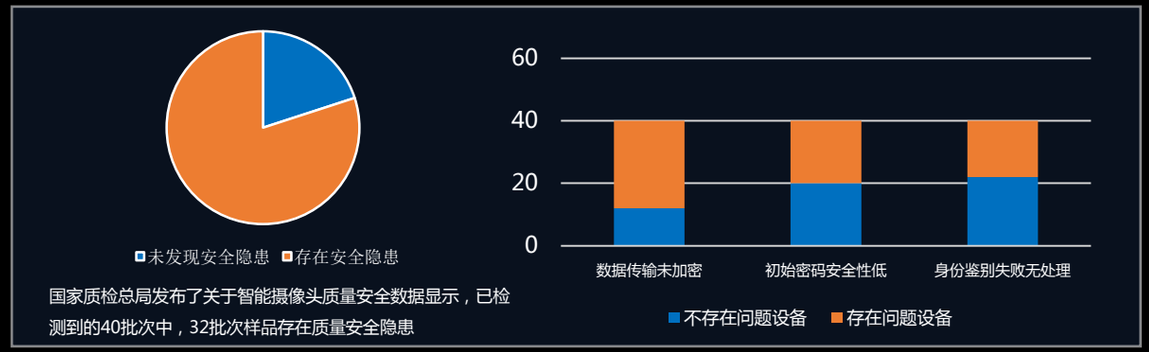
- AIoT生态安全现状不容乐观
- AIoT生态碎片化且产业链复杂
- AIoT安全的新仇旧恨

AIoT生态安全现状不容乐观



智能摄像头破解软件和破解的IP地址在黑市上频繁交易

百度安全专家利用设备系统漏洞，在无需物理接触、无需拆解门锁的情况下秒破某品牌互联网智能锁，获得开锁密码



国家质检总局发布了关于智能摄像头质量安全数据显示，已检测到的40批次中，32批次样品存在质量安全隐患

2017年智能设备安全典型案例

时间	物联智能设备	安全问题和漏洞	类型
4月22日	网络摄像头	360安全研究院披露http81IoT僵尸网络	DDoS攻击
5月9日	数十种物联网设备	绿盟科技发布《解读国内物联网资产暴露情况分析》	漏洞入侵
6月21日	DVR、路由器、摄像头等	Mirai利用类似蠕虫的方式感染	僵尸网络DDoS攻击
8月17日	iOS终端	iOS Secure Enclave（TEE系统）固件密钥被公布	漏洞入侵

AIoT生态安全现状不容乐观

泰尔终端实验室联合百度安全等机构厂商对12个知名品牌的智能电视进行安全评测，发现无一幸免存在安全问题。

模块	检测项	检测指标	存在风险的电视占比例
硬件	固件安全	是否存在固件更新风险；能否直接读取固件内容，还是仅能通过系统预留的更新接口来读写；固件符号调试信息是否去除	25.00%
操作系统	系统安全配置	系统是否进行了安全配置，如开启SElinux	91.67%
	调试安全	调试接口（adb）是否可远程非授权开启	75.00%
	安全启动	系统是否有安全启动校验机制，逐层验证系统的完整性，保证系统不被篡改	25.00%
	系统更新	是否有更新功能，更新过程是否有数据加密和身份验证，是否有完整性交验，是否有签名。	33.33%
	系统回滚	系统是否存在回滚风险，即是否不允许系统进行版本降级更新	41.67%
	安全漏洞检测	操作系统是否存在未修复已知高危或严重漏洞	100.00%
	端口安全	是否不存在开放端口，可供下载安装任意软件、远程静默安装、远程打开应用	75.00%
	第三方应用	是否关闭未知来源安装，不能允许随意安装第三方应用	66.67%
系统组件	控制权限管理	是否进行权限的校验；是否禁用工程模式	100.00%
	组件更新安全	关键组件更新是否采用签名校验	16.67%
	通信传输安全	系统通信是否采用安全传输信道，如SSL传输等	50.00%

模块	检测项	检测指标	存在风险的电视比例
系统组件	安全漏洞检测	系统组件是否存在已知漏洞	91.67%
	用户信息保护	是否存在未告知联网行为情况下传输用户数据	16.67%
	蓝牙协议	是否存在重放攻击等	33.33%
	内置应用组件	系统内置应用组件是否存在安全性风险，例如提权，信息泄露等问题	41.67%
控制APP	安全加固	是否经过加固保护	100.00%
	进程注入	是否有防注入防护功能	91.67%
	逆向分析	dex文件、SO文件能否被反编译进行逆向分析	100.00%
	二次打包	APK是否能够重打包	100.00%
	用户数据传输	数据传输是否存在未加密的用户关键数据	91.67%
预置第三方应用	安全漏洞检测	预置第三方应用是否存在中、高危安全漏洞	50.00%
	恶意行为检测	是否存在用户信息窃取、恶意吸费等恶意行为	33.33%
AI电视业务	语音模块	是否存在安全缺陷，如可窃取用户语音数据等	33.33%
	摄像头	是否存在被越权调用远程非授权开启	0.00%

AIoT生态碎片化严重

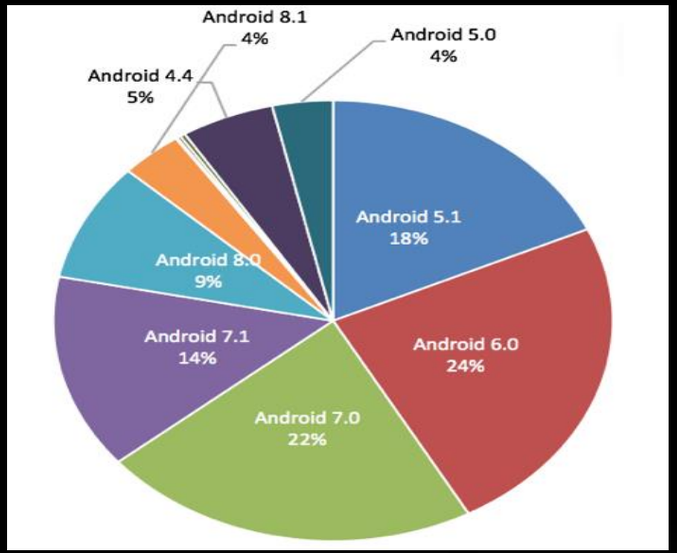
从PC到移动到AIoT，生态愈发碎片化，安全防护将面临更大挑战

产品形态种类
操作系统种类
芯片种类
终端厂商数量

AIoT生态产业链复杂

代码来源于产业链中的各个环节

终端成本低，设备软件系统版本普遍较旧，重要安全机制(SELinux、TEE等）缺失



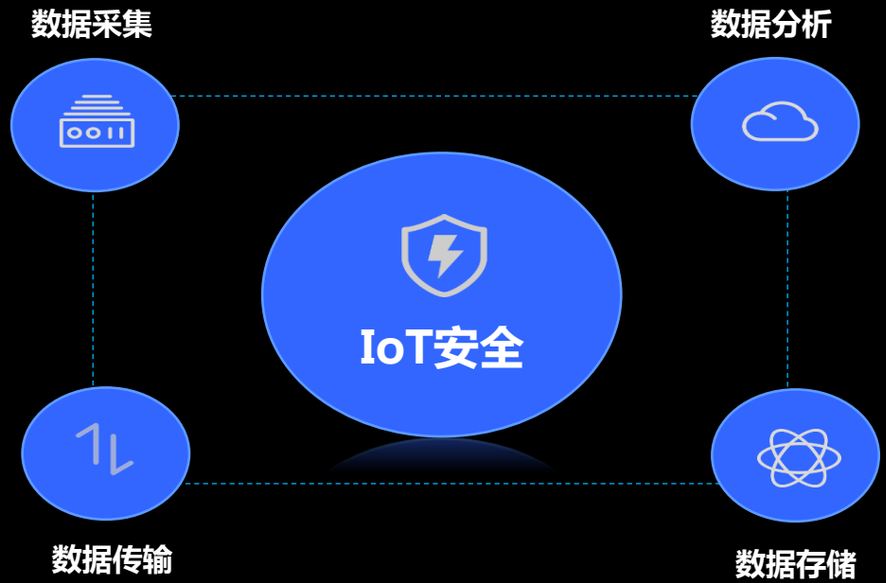
AIoT将面临系统性安全挑战

传感器欺骗	软件缺陷	数据风险	系统风险	网络风险
 摄像头	TensorFlow	 数据集	Linux	 劫持
 麦克风	Caffe	 对抗数据	Windows	 欺骗
 雷达	PaddlePaddle		Android	

AIoT安全的新仇旧恨

生物识别的马奇诺防线——破解人脸识别

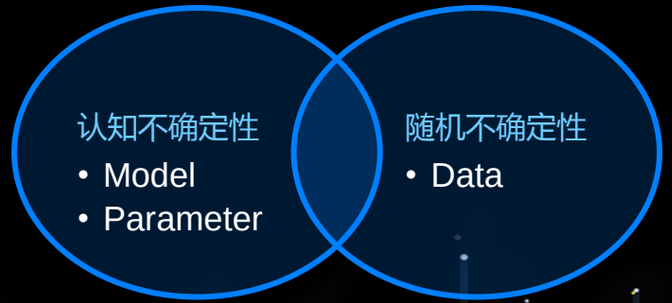
自动驾驶汽车将STOP“误读成”限速45英里/小时

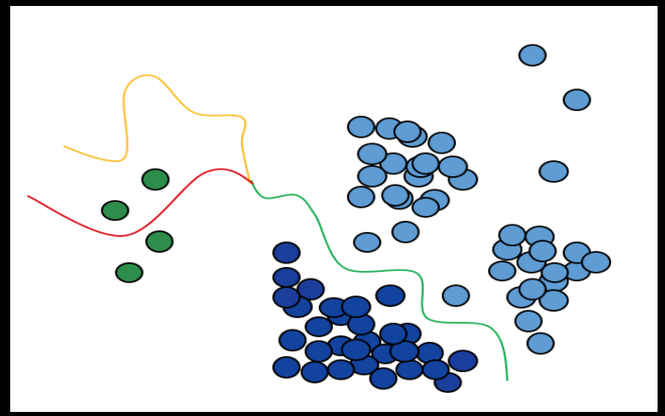


AI的核心安全挑战：不确定性

AI的不确定性

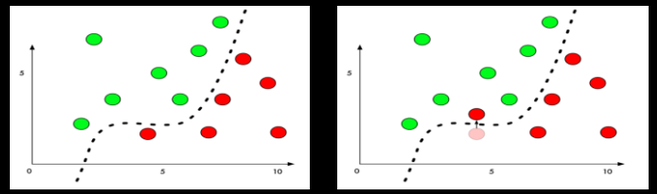
- 导致非预期/错误的输出
- 导致攻击者想要的预设输出

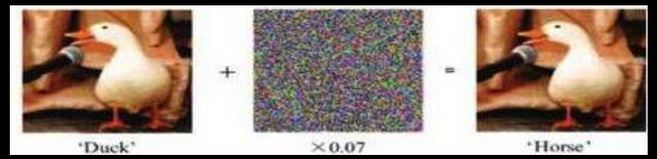


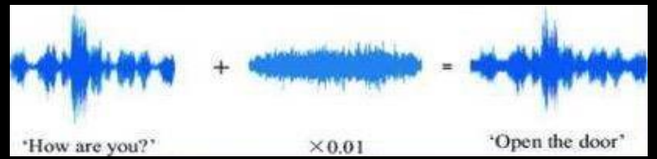


- 数据污染
- 数据流攻击
- 机器学习对抗性攻击

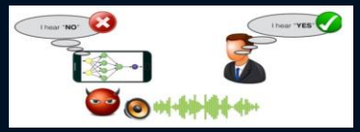
机器学习对抗性攻击



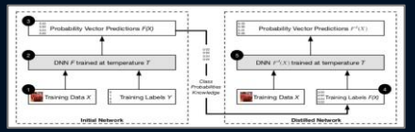




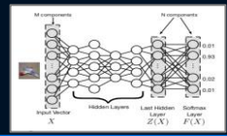
物理世界攻击
可以在现实生活中发起攻击



黑盒攻击
只要了解输入和输出



白盒攻击
需要了解模型内部细节

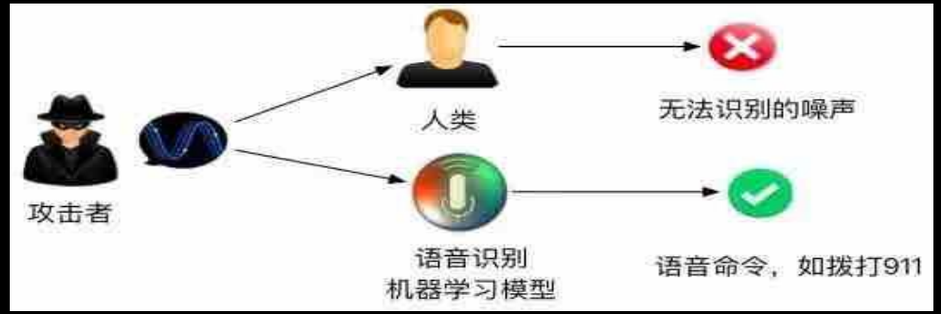


针对语音识别的机器学习对抗性攻击

隐匿语音命令攻击

攻击语音识别模型，使得人类难以识别的噪声，语音识别模型却可以识别为指定的命令，比如控制家居、播放音乐、购物以及转账等

可实现**物理世界攻击**，攻击过程只依赖空气传播，普通播音设备即可完成



攻击唤醒音

通过“隐匿语音命令攻击”唤醒智能音箱

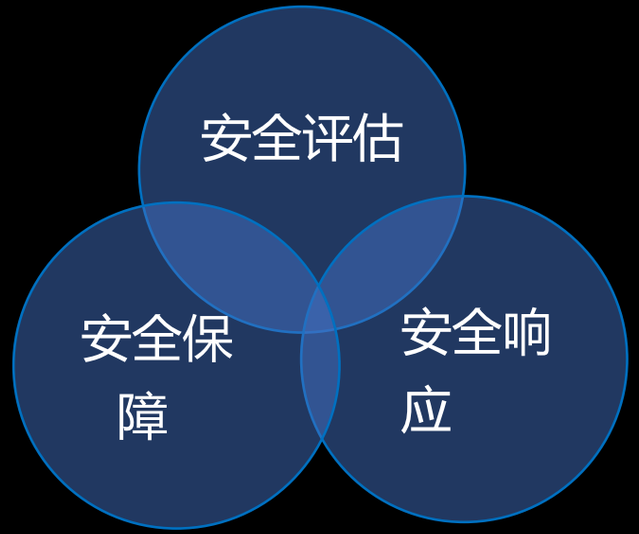
智能音箱在唤醒状态时，会实时上传周围的语音到云端
一旦唤醒音被攻破，这些设备就可能被攻击者利用成一个用户身边的**监听器**

AIoT安全问题解决思路

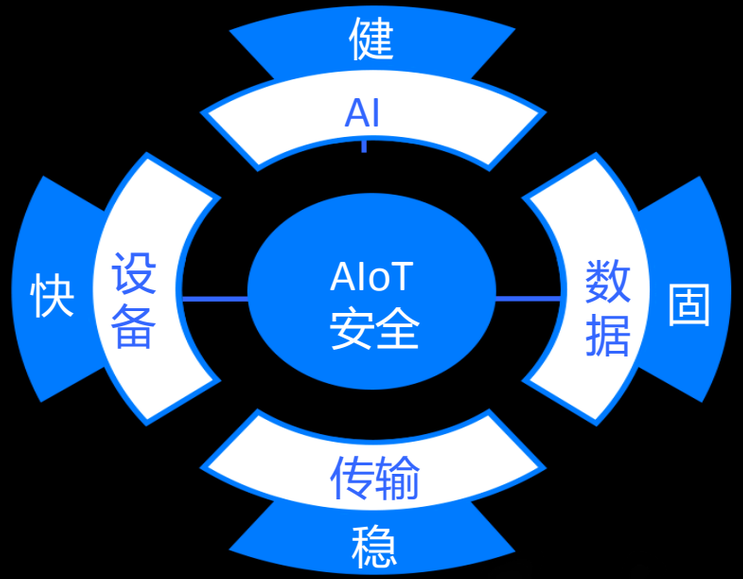
- 安全评估：建立标准防范于未然
- 安全保障：云管端整体防护系统
- 安全响应：问题修复能力是生命线

系统性解决AIoT的安全问题

智能设备架构图



安全评估：尽早发现安全隐患



对象	主要安全威胁	脆弱点
硬件	1. 非授权的访问； 2. 功能失效、设备不可用； 3. 假冒设备； 4. 重放攻击、侧信道攻击；	1. 信号注入 2. 传输未加密 3. 访问控制缺失
固件	1. 非授权的访问； 2. 审计数据丢失； 3. 恶意代码攻击； 4. DDoS攻击、溢出攻击、口令猜测、密码分析	1. 缺乏身份认证机制 2. 访问控制缺失 3. 调试接口暴露 4. 审计漏洞
应用软件	1. 非授权访问； 2. 软件漏洞； 3. 恶意代码攻击；	1. 弱口令 2. 缺少身份认证机制 3. 调试接口暴露
外围接口	1. 非授权访问； 2. 审计失效；	1. 缺少访问控制机制 2. 调试接口暴露 3. 审计漏洞
通信	1. 通信数据泄露、篡改、丢失； 2. 传输中断、拦截、篡改、伪造； 3. 拒绝服务攻击，重放攻击，中间人攻击； 4. 虚假路由； 5. 通信协议漏洞	1. 采用明文传输 2. 通信协议存在漏洞 3. 安全协议存在漏洞
用户数据	1. 用户数据泄露	1. 终端过度手机 2. 传输、存储安全漏洞

安全评估：利用检测工具自检自查

Advbox

对抗样本工具包

- 实现多种生成对抗样本的攻击方法，包括FGSM、BIM、DeepFool、JSMA

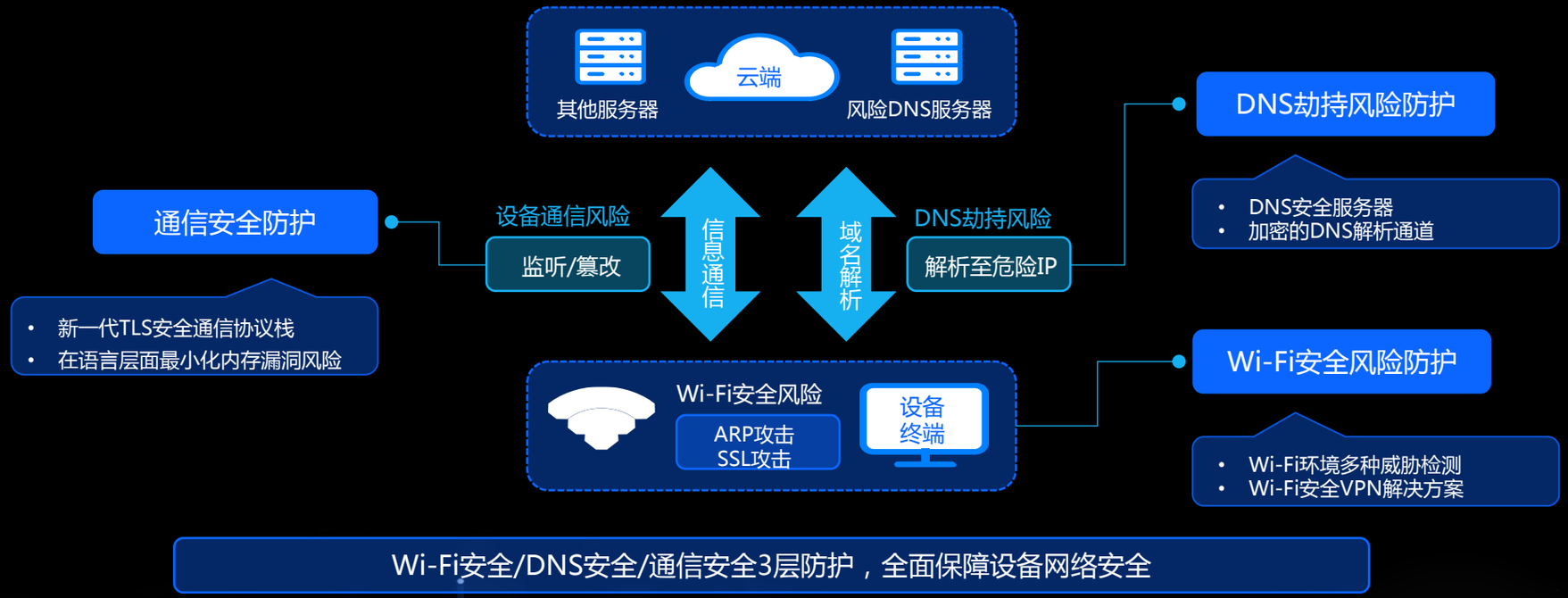
百度锐眼

IoT安全检测工具

- 专门为智能设备打造的安全检测工具，可检测系统漏洞、恶意程序及病毒、SELinux配置

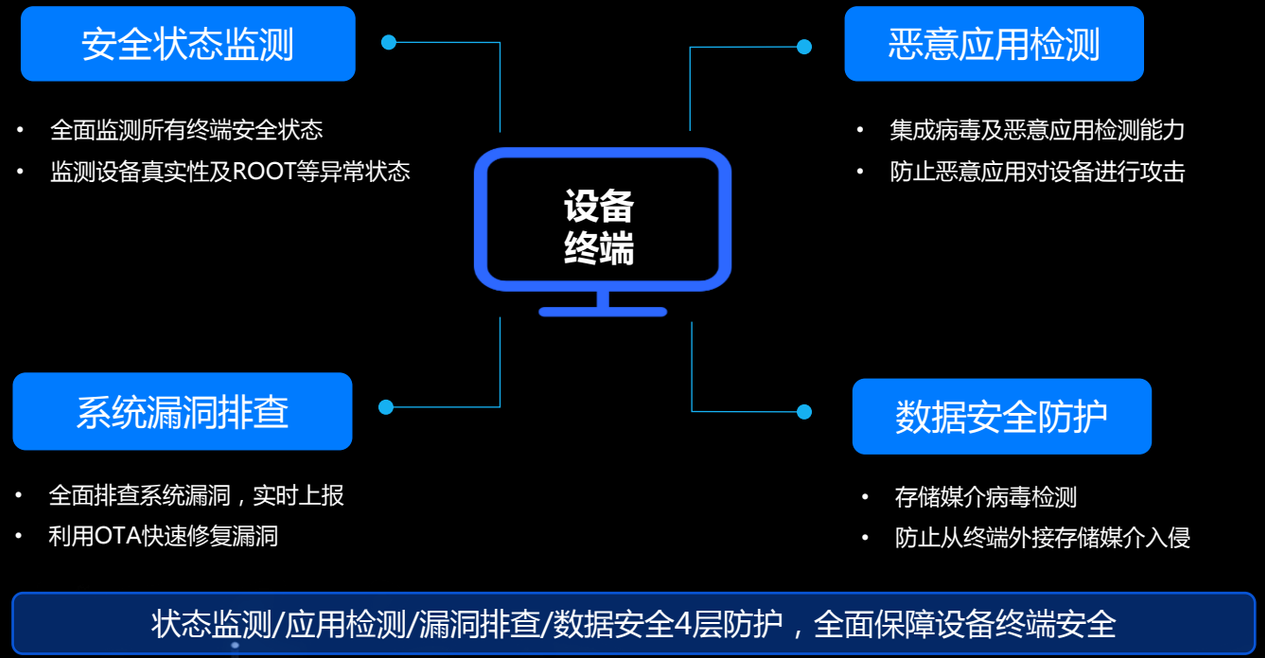
锐眼官网：ruiyan.baidu.com

安全保障：云端防护思路



Wi-Fi安全/DNS安全/通信安全3层防护，全面保障设备网络安全

安全保障：设备终端防护思路



状态监测/应用检测/漏洞排查/数据安全4层防护，全面保障设备终端安全

安全响应：建立升级及修复通道

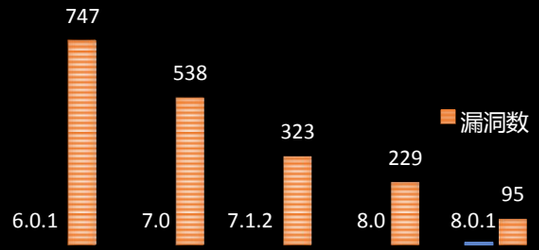
修复已知问题

- ✓ 漏洞补丁
- ✓ 修复bug

加强安全防护

Android系统版本越高，安全手段与漏洞缓解措施越强

系统版本越高，漏洞风险越低



Google官方公布的版本漏洞分布数量

系统版本越高，安全配置越强

- SELinux安全机制（强制访问控制系统）从4.4版本的44个到7.0版本的0个
- 8.0版本之后用户必须授予权限，才能从不是第一方应用商店的来源安装应用，能避免用户在不知情的情况下被安装恶意应用
- 7.0版本加强了文件加密，可以更好地隔离和保护设备上的不同用户和资料

安全响应：关注升级通道本身的安全性

百度安全 OTA

IoT固件升级服务

- 兼容Android、Linux及嵌入式系统
- 通过集成多种安全能力保障升级过程无虞

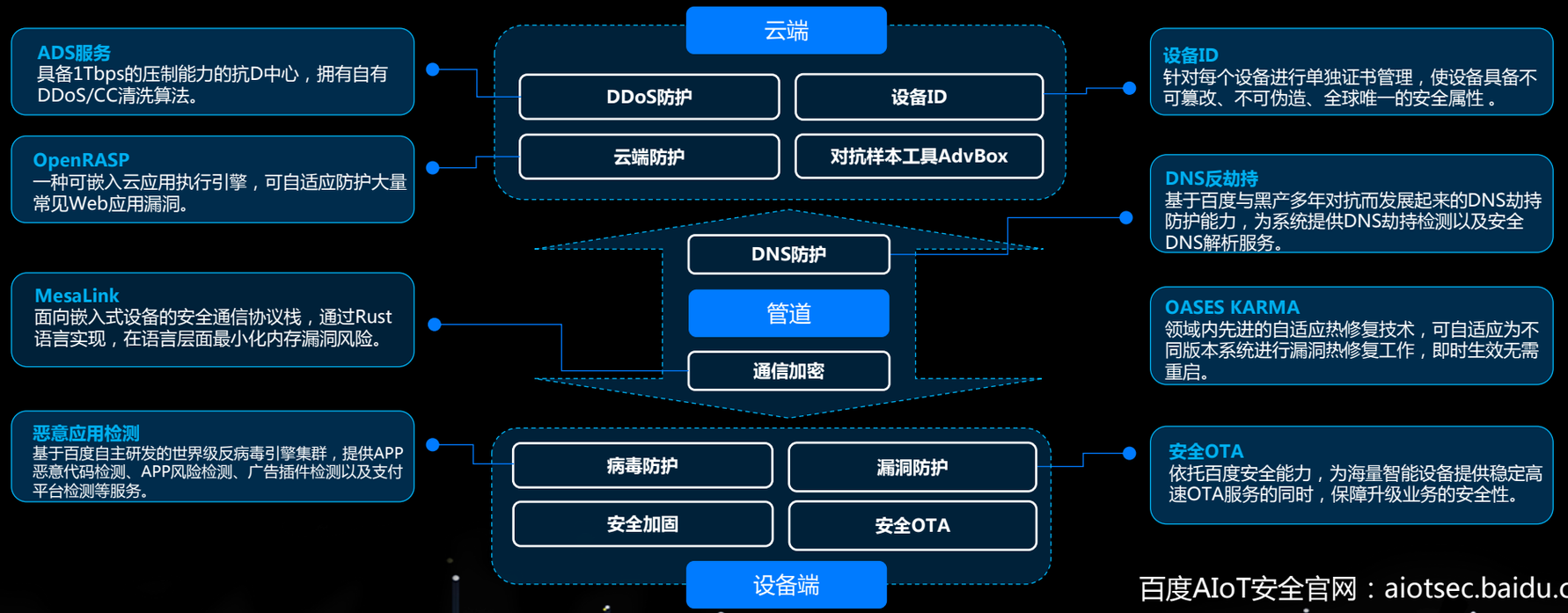
百度安全OTA官网：ota.baidu.com

OASES KARMA

Android系统自适应热修复漏洞能力

- 漏洞补丁自适应不同系统，免去针对不同系统打不同补丁的麻烦
- 漏洞热修复，无需重启机器，修复漏洞过程不影响设备使用

百度安全AIoT安全解决方案



百度AIoT安全官网：aiotsec.baidu.com

THANK YOU

合作联系：aiotsec@baidu.com