安全智能数据收集探索实践

韩广利

TalkingData 研发总监

Agenda

移动端发展趋势

Android / iOS
发展趋势

数据安全

数据处理流程
国内外数据信息安全规范要求
移动端数据存储传输数据安全实践

总结展望

能收集什么

理论上

- OS层面允许的都可以
  - ✓ 获取权限范围内
- 与一个App 可以覆盖的功能一致

事实上

- 不能为所欲为，仅按需使用
  - ✓ 继承和基于App 的权限和用户授权
- 被动收集，非主动采集

SDK有哪些类别

SDK 功能分类

- 基础能力类：

  基础统计、使用统计、推送、语音、视频、认证（短信、人脸）、聚合支付

- 场景融合类：

  广告监测、聊天室、视频直播、社区、广告推广、推荐、防作弊、姿态识别

持续更新的TalkingData SDK

iOS

2011/10/09
MRC->ARC

2012/09/01
Data
Privacy
idfa/idfv

2013/09/09
64bit
CoreMotion
CoreLocation

2014/09/12
iWatch
Apple Pay

2015/09/09
ATS/Bitcode/Swift

2016/06/01
IPv6

2017/09/20
AI/AR

2018/09/18
Dual Sim

2019/09/20
Dark Mode
Security
Performance

Android

2010/12/06
Android ID

2011/10/18
SerialNo

2012/11/13
TLSv1.2

2014/11/12
Dual Sim
Wear

2015/10/05
Http Client
Permissions

2016/08/22
Security
Permissions

2017/08/22
TensorFlowLite

2018/08/07
AI/DeepMind

2019/09/05
Dark Mode
Security
Performance

2011 2012 2013 2014 2015 2016 2017 2018 2019

TalkingData

20 +种主流平台支持

Unity

Cordova

H5

Cocos

PhoneGap

小程序

Lua

Ionic

快应用
Quick App

FlashAir

Hybrid

Flutter

APICloud

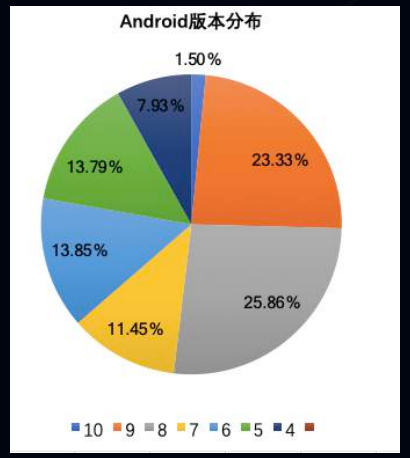
React native

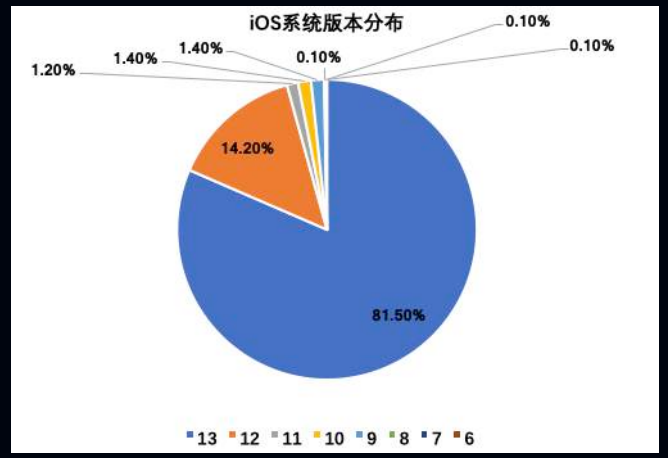
Android vs. iOS

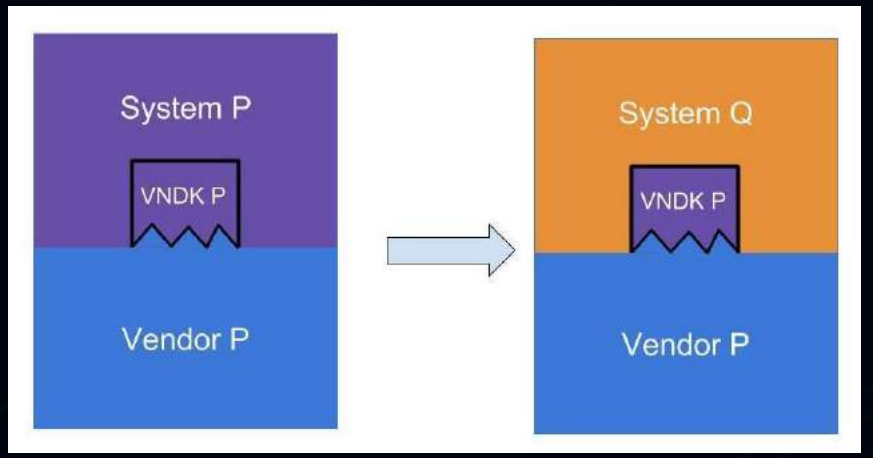
Android和iOS设备数量比例？

Android/iOS系统演进

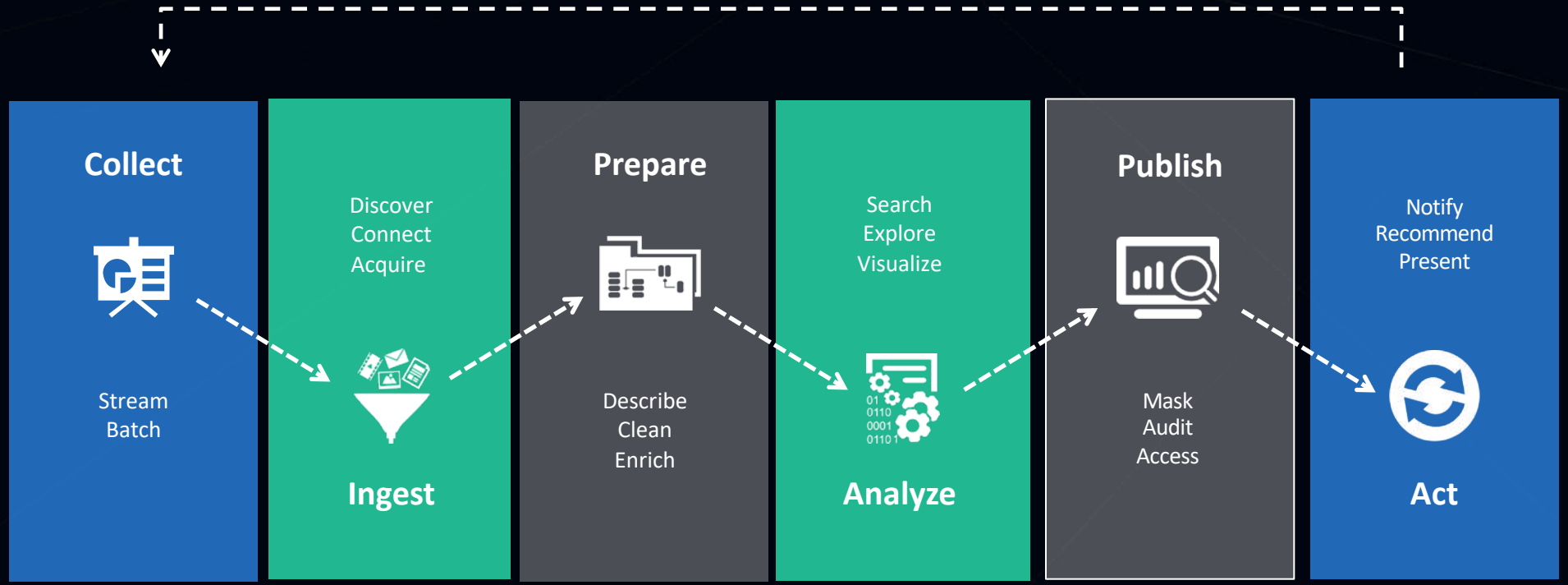
Android碎片化



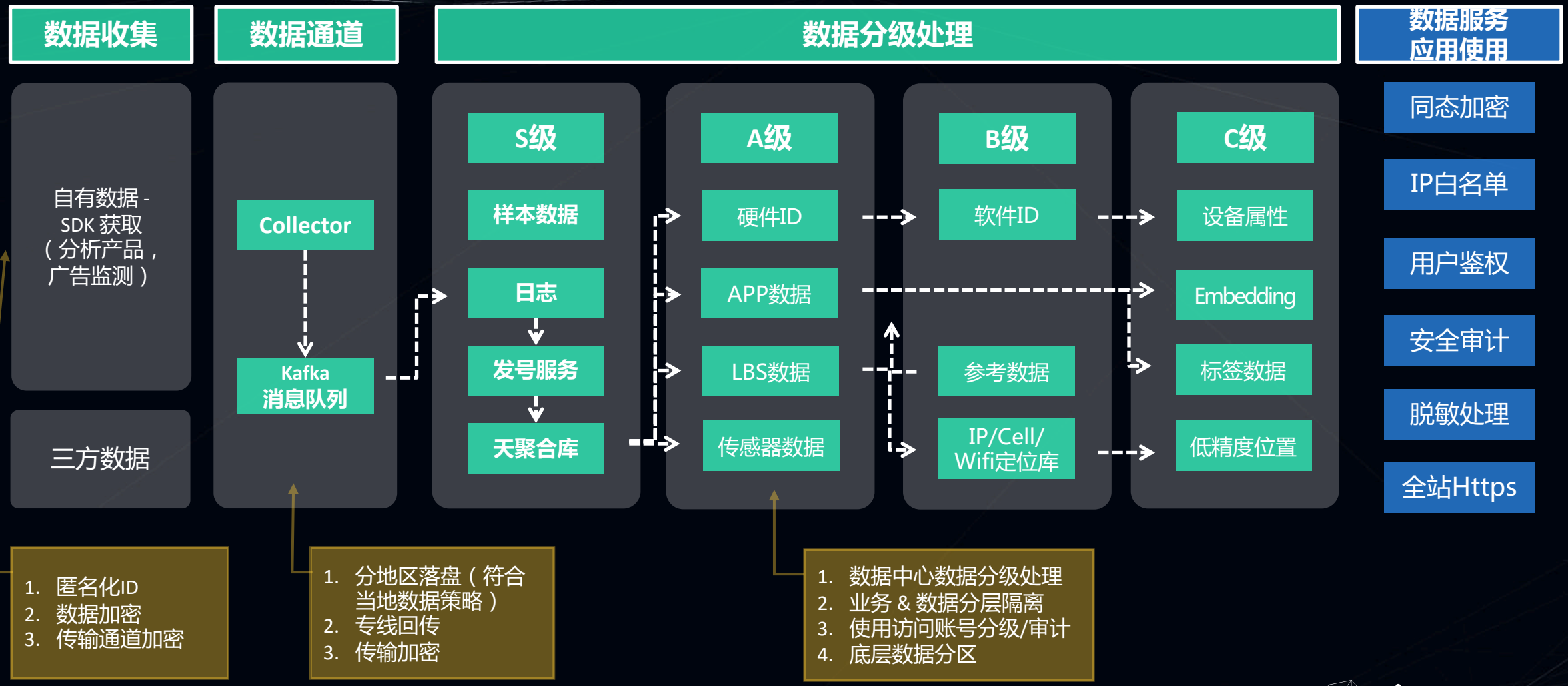




数据处理流程


Collect
Stream
Batch
Discover
Connect
Acquire
Ingest
Prepare
Describe
Clean
Enrich
Search
Explore
Visualize
Analyze
Publish
Mask
Audit
Access
Notify
Recommend
Present
Act

数据处理流程中安全管控



TalkingData

*参考数据：APP标准化，机型标准化，网络标准化

国内外数据信息安全的规范

Google Play 开发者规范

- https://play.google.com/about/developer-content-policy
- https://play.google.com/about/privacy-security-deception/
- https://developer.android.google.cn

Apple Store 开发者规范

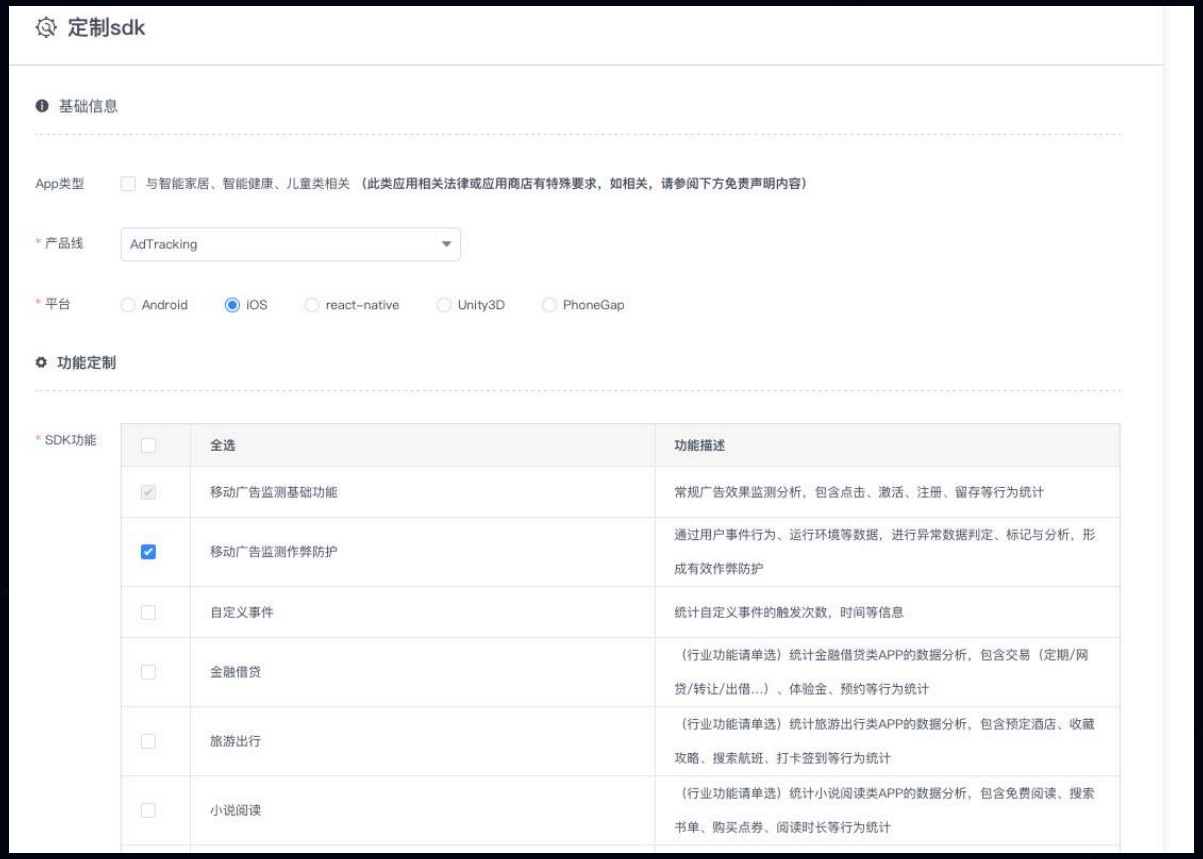
- https://developer.apple.com/app-store/review/guidelines/

各国/地区数据 & 隐私保护的法律规范

- 中国 个人信息安全规范： https://www.tc260.org.cn
- 欧盟 GDPR： https://eugdpr.org
- 美国 COPPA： http://www.coppa.org

SDK 按需定制 - 开发者自主选择，功能&权限匹配

- https://www.talkingdata.com/spa/sdk/#/config?productLine=AppAnalytics&sdkPlatform=Android



SDK 按需定制 - 开发者自主选择，功能&权限匹配

- https://www.talkingdata.com/spa/sdk/#/config?productLine=AppAnalytics&sdkPlatform=Android

您知悉并同意，为提升服务质量、用户体验以及其他实现【请根据实际需要补充】等服务功能之必要目的，我们会在必要范围内收集、存储、加工或处理您的个人信息或您在使用服务中形成的数据信息。

我们有权选择使用第三方合作伙伴提供的数据统计分析服务（例如：TalkingData统计分析SDK等），并由合作方收集、使用、加工和处理个人信息。

TalkingData

移动端数据存储加密实践：动态秘钥分存机制

例子：A,B,C三人创立了一家公司。三人把公司的机密信息放在保险柜里。为了实现共同决策，三人想设计一种密码方案，只有三人同时到场才能开启保险柜。于是三人找到了科学家S，并提出需求：

加密过程:

S把密码设定成了 ***666***

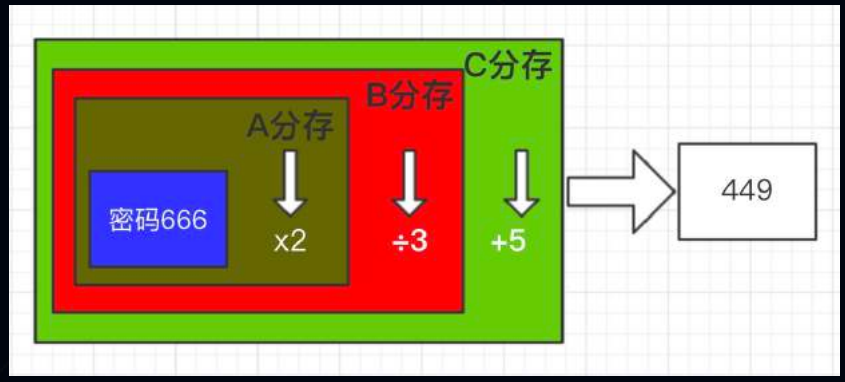
先对666进行了 ***x2*** 操作，得到了 ***1332***

再对1332进行了 ***÷3*** 操作，得到了 ***444***

最后在对444进行了 ***+5*** 操作，最后得到了 ***449***

S把明文密码 ***666*** 销毁，把密文 ***449*** 给到三人，

并将 ***+5*** 操作， ***÷3*** 操作，***x2*** 操作分别作为密码给C,B,A三人



解密操作:

每次想开启保险柜，三人必须将手中分存的秘钥 ***+5*** 操作， ***÷3*** 操作，***x2*** 操作

连同密文 ***[449]*** 一同交给S

由S先对密文 ***449*** 用A的 ***+5*** 进行逆操作也就是 ***-5***，得到 ***444***，

再对444用B的秘钥进行 ***÷3*** 逆操作也就是 ***x3***，得到 ***1332***，

最后再用C秘钥 ***x2*** 对 ***1332*** 进行逆操作也就是 ***÷2***，得到最后的密码 ***666***

移动端数据存储加密实践：动态秘钥分存机制

其它方案？

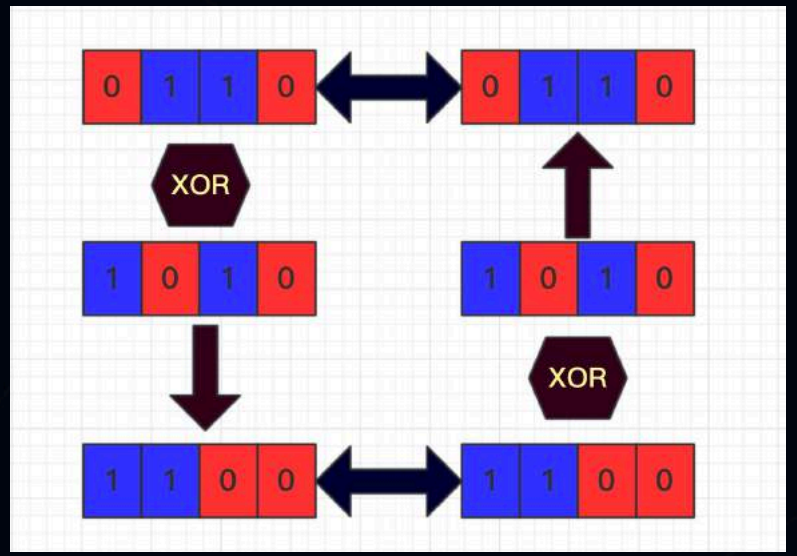
一个单位对另一个基本单位(0或1)连续进行两次异或操作之后，得到的结果就是是这个单位本身。也就是说**异或运算的逆运算是它本身**

加密：

- 同样设定明文密码 ***666***
- 将 ***666*** 对 ***2,3,5*** (这里2，3，5只是随意制定，理论上可以制定跟秘钥二进制位数相等任意字符)进行异或操作，得到密文***E***。
- 将 ***2,3,5*** 作为分存秘钥分发给 ***A,B,C*** 。并将密文 ***E*** 公开，将密码 ***666*** 销毁。

解密：

- 只需要将密文 ***E*** 和秘钥 ***2,3,5*** 交给 ***S*** 。***S*** 对密文 ***E*** 进行 ***2,3,5*** 的异或操作，便可“抵消”之前的操作，从而得到密码 ***666***



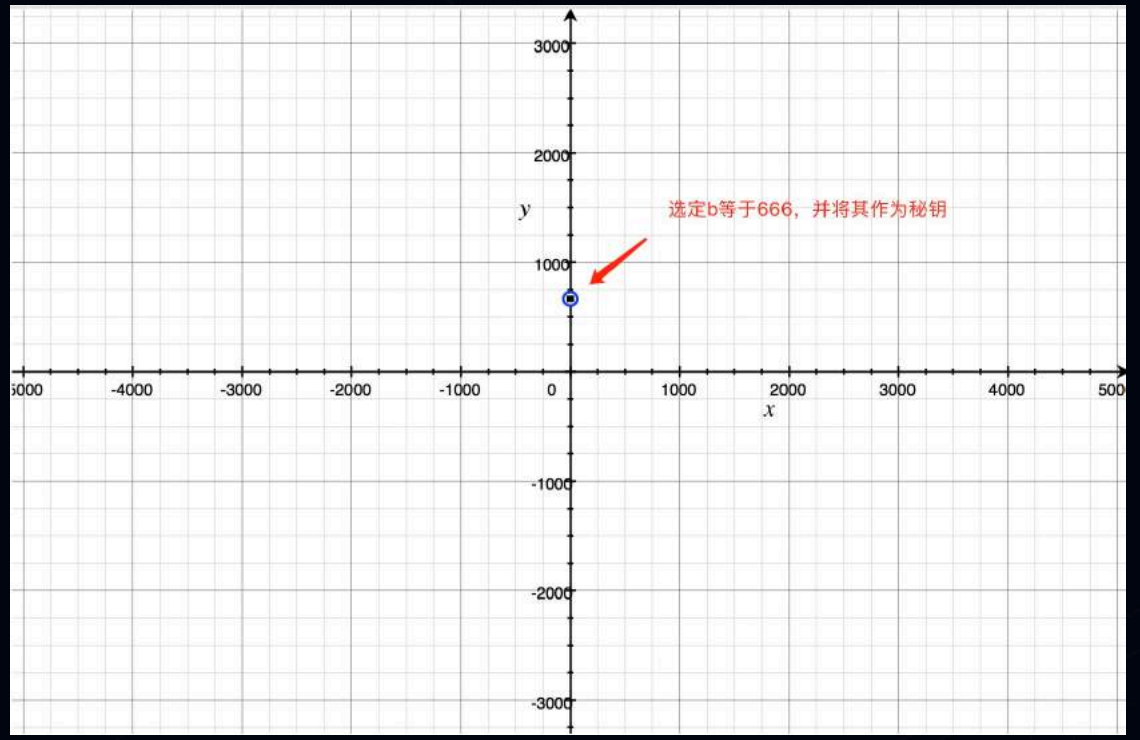
移动端数据存储加密实践：动态秘钥分存机制

例子：由于公司发展，A,B,C三人公务繁多，很难同时凑到一起。于是三人再次找到S，想获取一种方案，这种方案中三人只要两人同时到场，便可开启保险柜。S应该怎么办？

方案：$y=kx+b$

加密：

- S同样把明文密码设定成 666 ，并将其赋值给 b



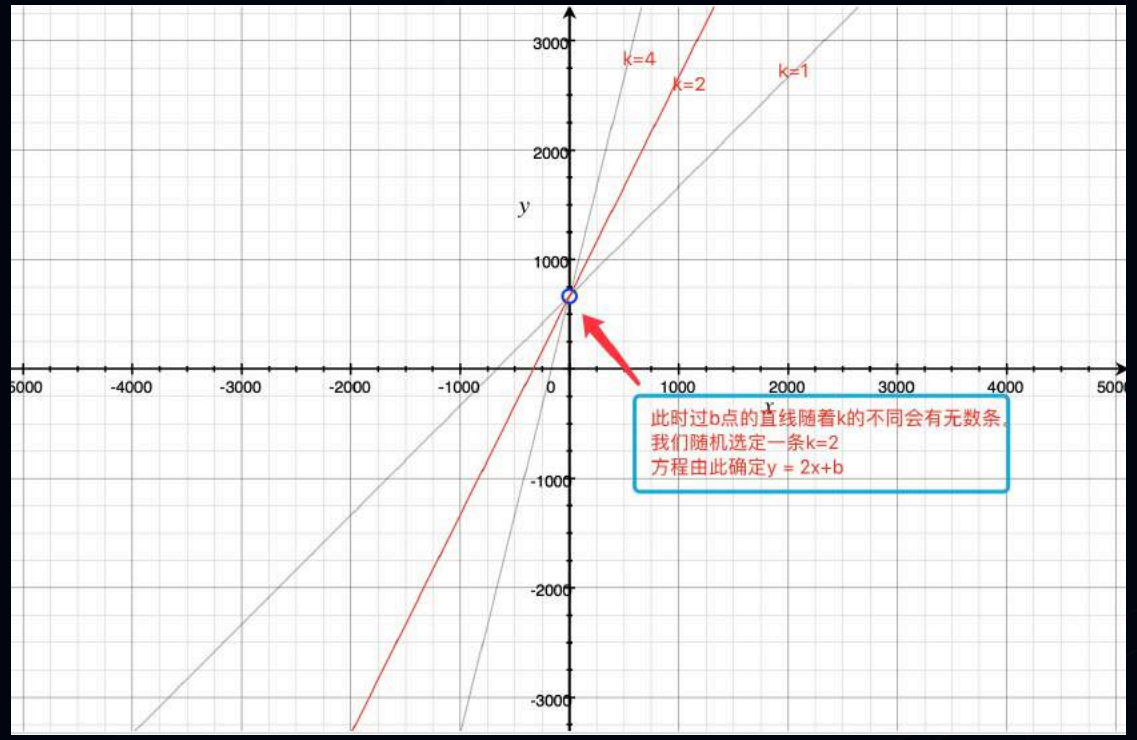
移动端数据存储加密实践：动态秘钥分存机制

例子：由于公司发展，A,B,C三人公务繁多，很难同时凑到一起。于是三人再次找到S，想获取一种方案，这种方案中三人只要两人同时到场，便可开启保险柜。S应该怎么办？

方案：$y=kx+b$

加密：

- S同样把明文密码设定成 666 ，并将其赋值给 b
- 随机选定了一个 k = 2 ,所以等式就变成了
- y = 2x + 666



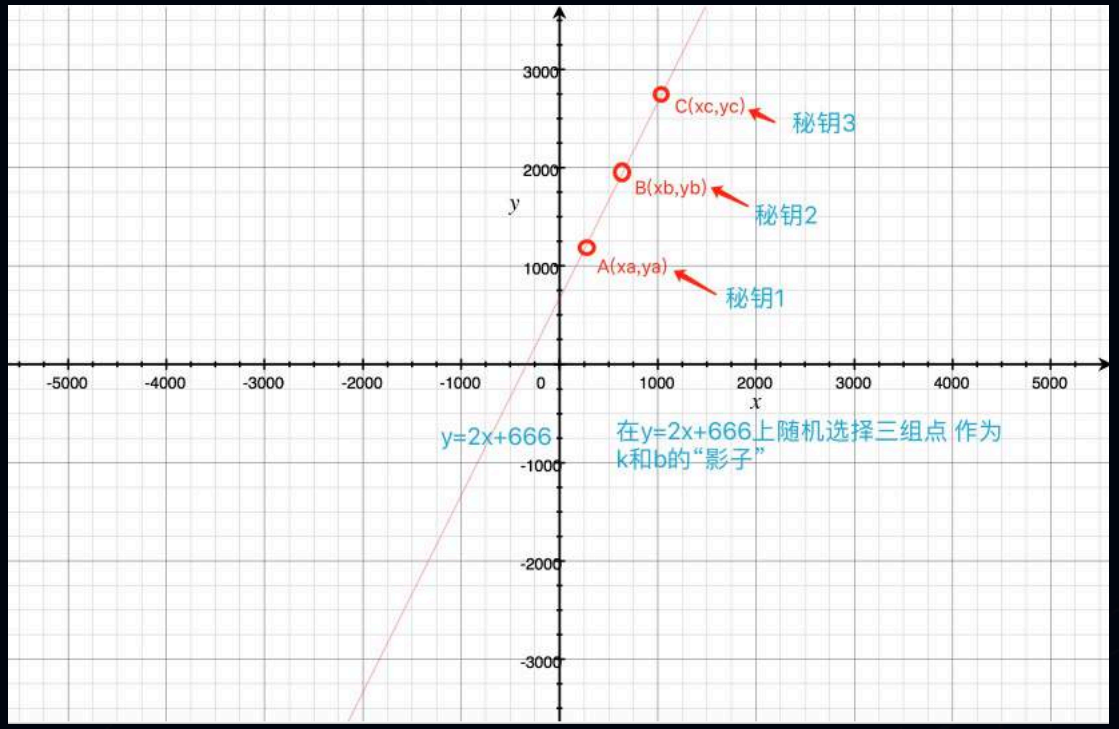
移动端数据存储加密实践：动态秘钥分存机制

例子：由于公司发展，A,B,C三人公务繁多，很难同时凑到一起。于是三人再次找到S，想获取一种方案，这种方案中三人只要两人同时到场，便可开启保险柜。S应该怎么办？

方案：$y=kx+b$

加密：

- S同样把明文密码设定成 666 ，并将其赋值给 b
- 随机选定了一个 k = 2 ,所以等式就变成了
- y = 2x + 666
- S分别使x=1,2,3..... 得到对应的 y =668,670,672.....
- 把(x=1,y=668),(x=2,y=670),(x=3,y=672)分别作为秘钥分发给A,B,C三人。



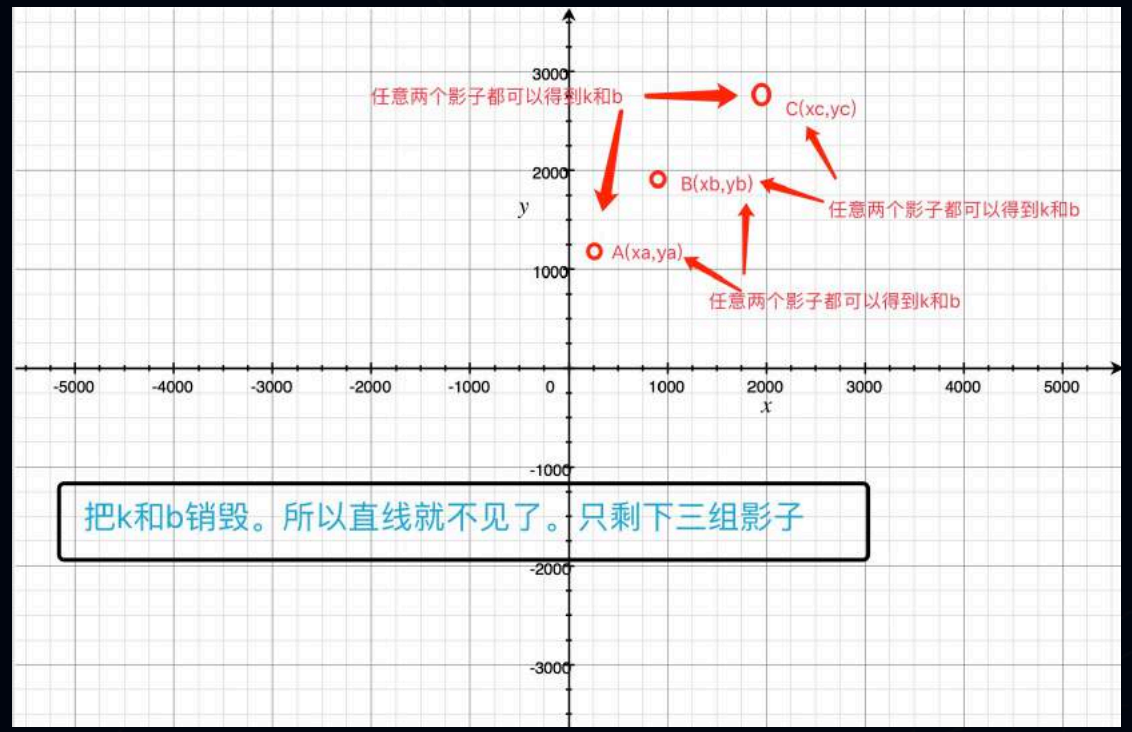
移动端数据存储加密实践：动态秘钥分存机制

例子：由于公司发展，A,B,C三人公务繁多，很难同时凑到一起。于是三人再次找到S，想获取一种方案，这种方案中三人只要两人同时到场，便可开启保险柜。S应该怎么办？

方案：$y=kx+b$

加密：

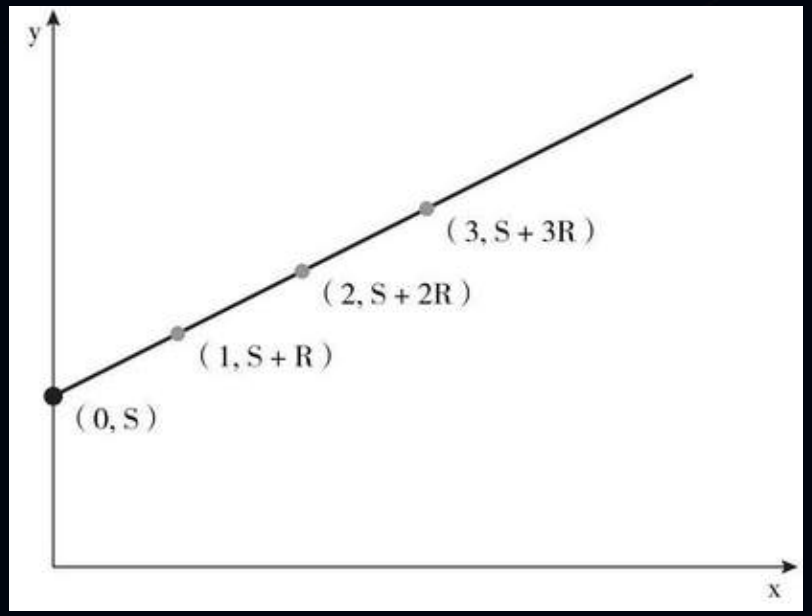
- S同样把明文密码设定成 666 ，并将其赋值给 b
- 随机选定了一个 k = 2 ,所以等式就变成了
- y = 2x + 666
- S分别使x=1,2,3..... 得到对应的 y =668,670,672.....
- 把(x=1,y=668),(x=2,y=670),(x=3,y=672)分别作为秘钥分发给A,B,C三人。
- 销毁k，销毁b。



TalkingData

移动端数据存储加密实践：动态秘钥分存机制

密钥分存：
密钥被分成N个片段，只要我们获得其中的K个片段(子密钥)，就可以把原密钥重新还原。但如果获得的片段数量少于K，就无法知道关于密钥的任何信息



密钥分存的几何示例（N=2）

S：原始秘钥

R：子秘钥1

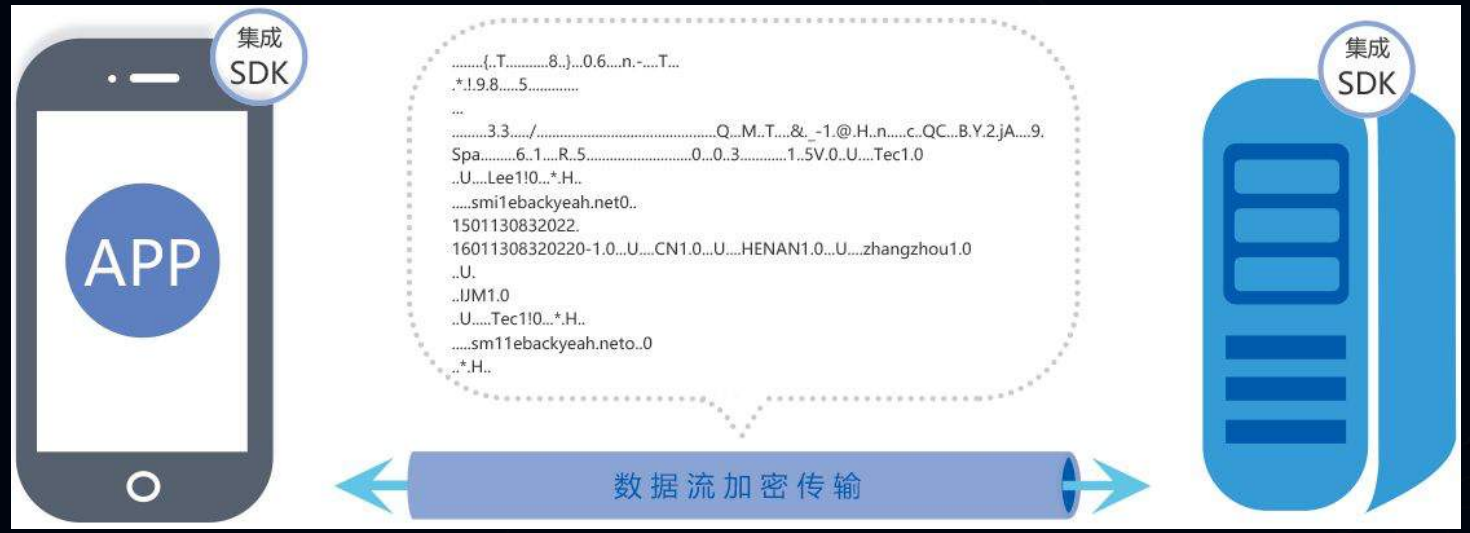
S⊕R：子秘钥2

⊕代表逻辑算符互斥

移动端数据传输加解密实践

被动收集，非主动采集

- 开发者将SDK集成到APP中，添加SDK启动代码和埋点代码
- APP启动后，调用SDK启动接口，SDK启动
- APP退出或者系统回收后，SDK也相应退出和被回收
- 当触发到APP埋点，SDK接收到埋点数据，数据本地进行加密处理
- SDK数据通过https协议进行数据发送，数据报文加密处理



移动端数据传输加解密实践

算法选择（从性能和安全性综合）

- 对称加密AES
- 非对称加密: ECC\RSA
- 消息摘要: MD5
- 数字签名:DSA
- 轻量级：TEA、RC系列（RC4），Blowfish （不常换密钥）

* 速度排名：IDEA < DES < GASTI28 < GOST < AES < RC4 < TEA < Blowfish

名称	数据大小（MB）	时间（S）	平均速度MB/S	评价
DES	256	10.5	22.5	低
3DES	256	12	12	低
AES(256-bit)	256	5	51.2	中
Blowfish	256	3.7	64	高

移动端数据传输加解密实践

- 压测数据量：20W条
- 数据大小：原始70K+，压缩45K+
- 数据总大小：13G+
- 运行模式：4进程
- CPU：50%
- 内存：25%

加密类型	单条数据（压缩）	单条数据（压缩+加密）	解密耗时（分钟）	QPS
无	45K+	45K+	38	88
RC4	45K+	45K+	40	83
RSA	45K+	50K+	385	9

移动端数据传输加密实践

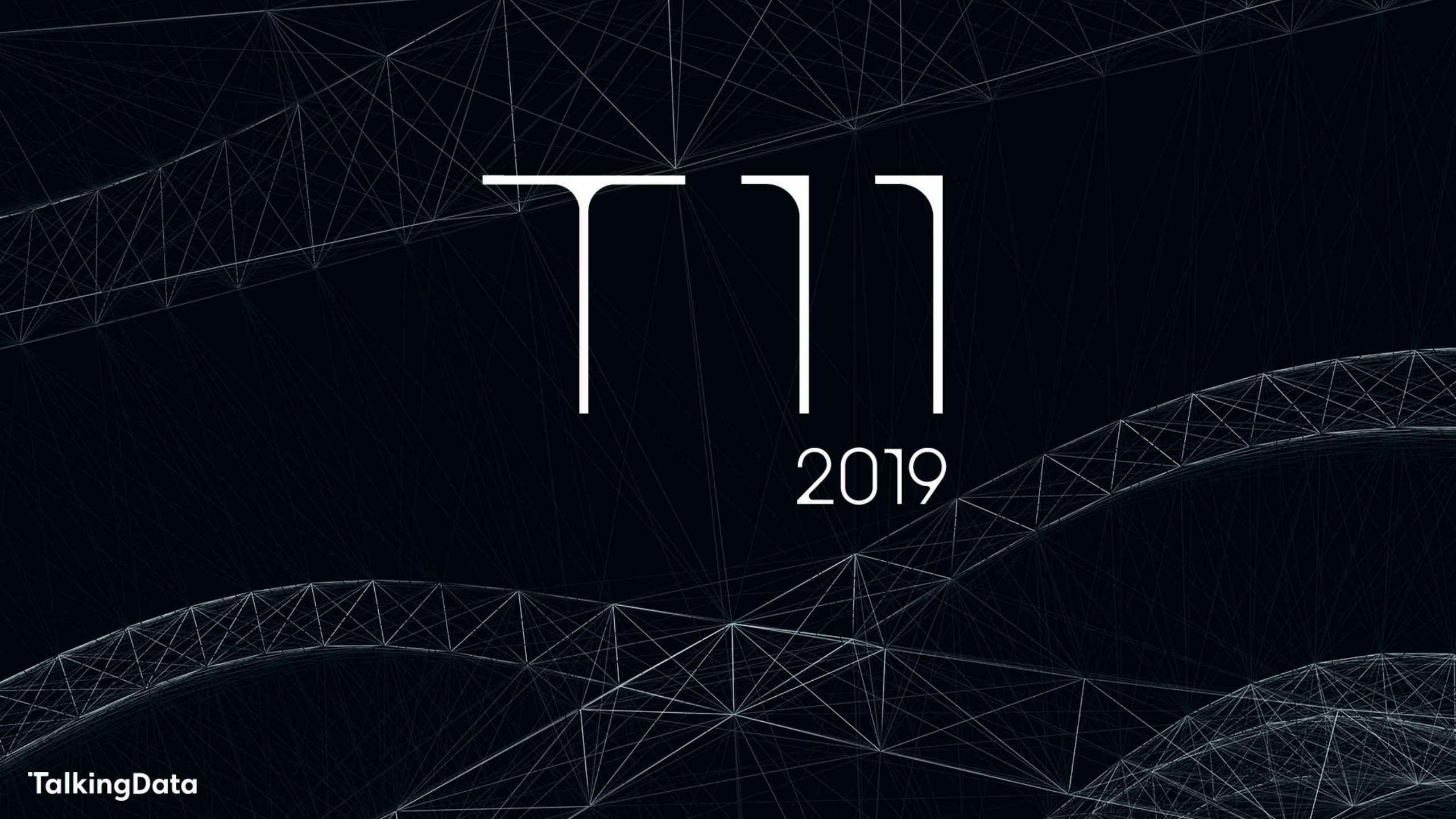
单钥密码算法性能比较

名称	实现方式	运算速度	安 全 性	改进措施	应用场合
DES	40-56bit 密钥	一般	完全依赖密钥，易受穷举搜索法攻击	双重、三重DES，AES	适用于硬件实现
IDEA	128bit密钥 8轮迭代	较慢	军事级，可抗差值分析和相关分析	加长字长为32bit、密钥为256bit， 采用232 模加、232+1模乘	适用于ASIC设计
GOST	256bit密钥 32轮迭代	较快	军事级	加大迭代轮数	S盒可随机秘 密选择，便于软件实现
Blowfish	256-448bit 密钥、16轮迭代	最快	军事级、可通过改变密钥长度调整安全性		适合固定密钥场合，不适合常换密钥和智能卡
RC4	密钥长度可变	快DES l0倍	对差分攻击和线性攻击具有免疫能力， 高度非线性	密钥长度放宽到64bit	算法简单，易于编程实现
RC5	密钥长度和迭代轮数均可变	速度可根据三个参数的值进行选择	六轮以上时即可抗线性攻击、通过调整字长、密钥长度和迭代轮数可以在安全性和速度上取得折中	引入数据相倚转	适用于不同字长的微处理器
CASTl28	密钥长度可变、16轮迭代	较快	可抵抗线性和差分攻击	增加密钥长度、形成CAST256	适用于PC机和UNIX工作站

用 数据+科技 的能力为客户创造价值

100个+	7.5亿+部	45万+款	30万+	1000家+
连接渠道	月活终端设备	服务移动应用	服务开发者	服务企业

TalkingData


2019
TalkingData