PyCon China 2020

PyConChina 2020
PYTHON 中国开发者大会 2020

Python For Good

《使用Python打造轻量级APP自动审计平台》

蔺国程

目录

Contents

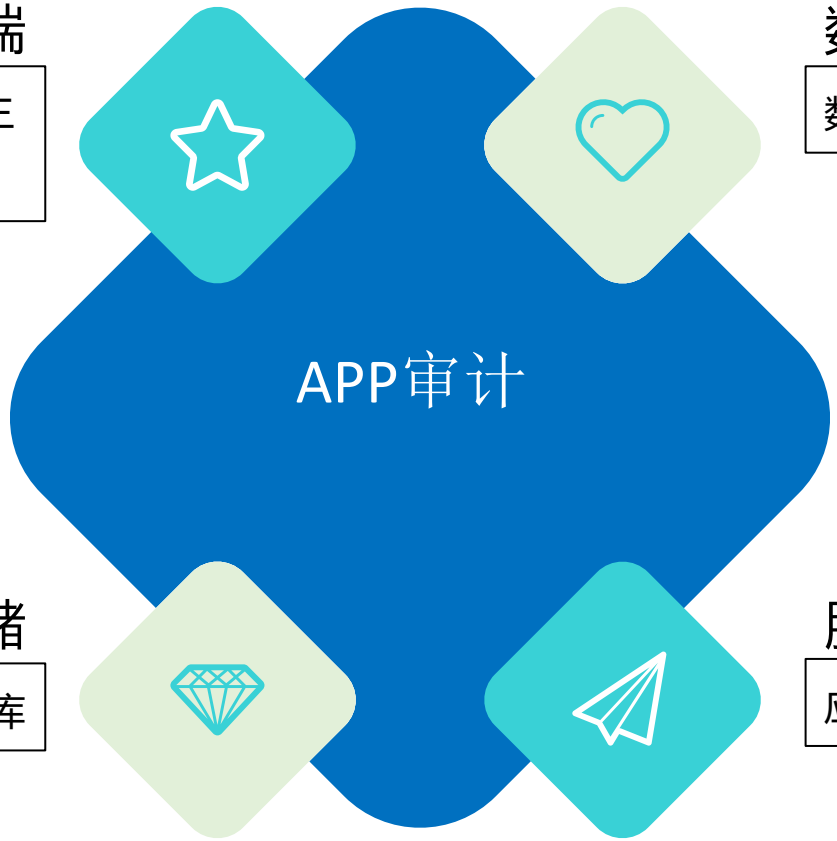
宏观视角下的APP审计

客户端

客户端完整性、组件安全、调试、注入、第三方库

数据传输

数据窃听、数据篡改、证书校验、隐私权限



数据存储

日志、数据库

服务端

应用层拒绝服务、流量解密、常规WEB渗透

自动化审计诉求

自动化审计

=

APP审计

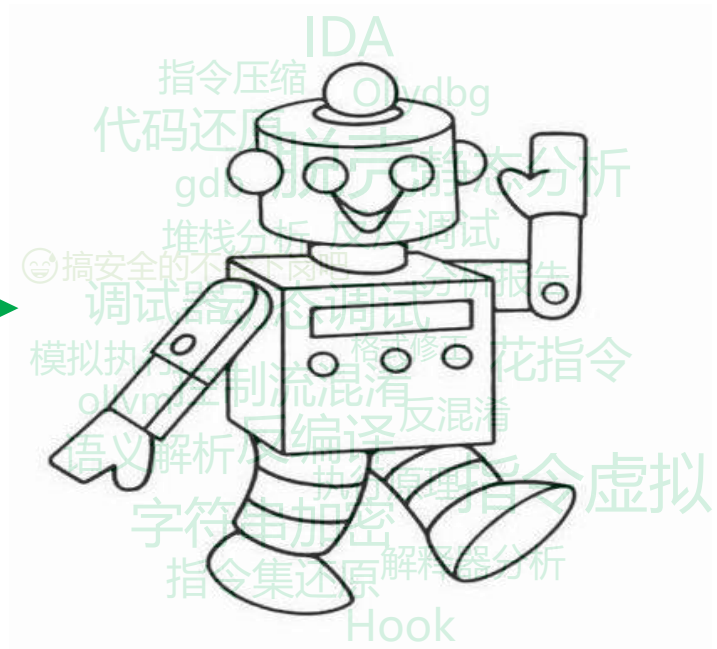
+

自动化

自动化审计

自动化	定向分析	可拓展	数据处理
字符串	关键字	规则拓展	采集
URL	特征提取	特征拓展	过滤
关联性	符号分析	威胁情报拓展	记录
加壳识别	系统调用表	模型拓展	回溯
算法报告	关联性分析	黑白名单	
协议报告	数据依赖分析		
威胁报告	隐私权限		
	第三方库		

理想情况



设计方案

源码

核心算法

机密信息

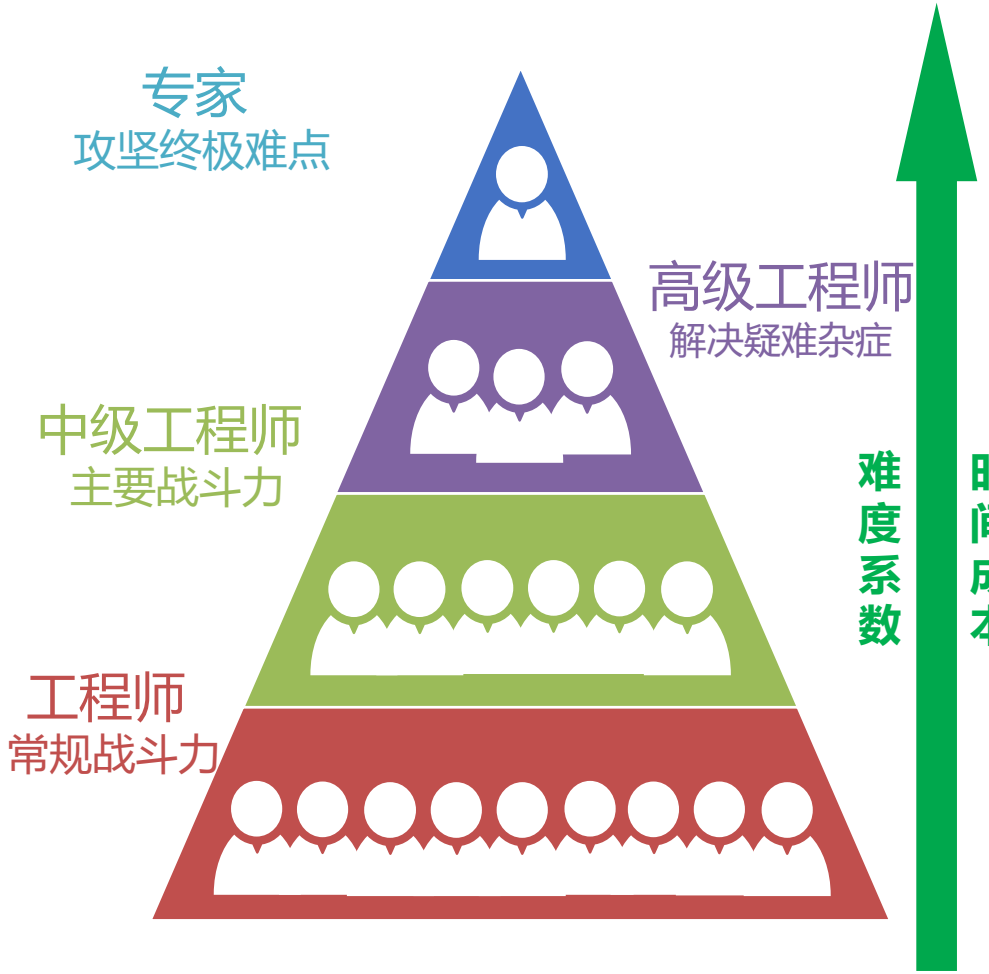
各种情报

0Dday

渗透教程

...

实际情况



大部是耗时耗力的重复机械性工作

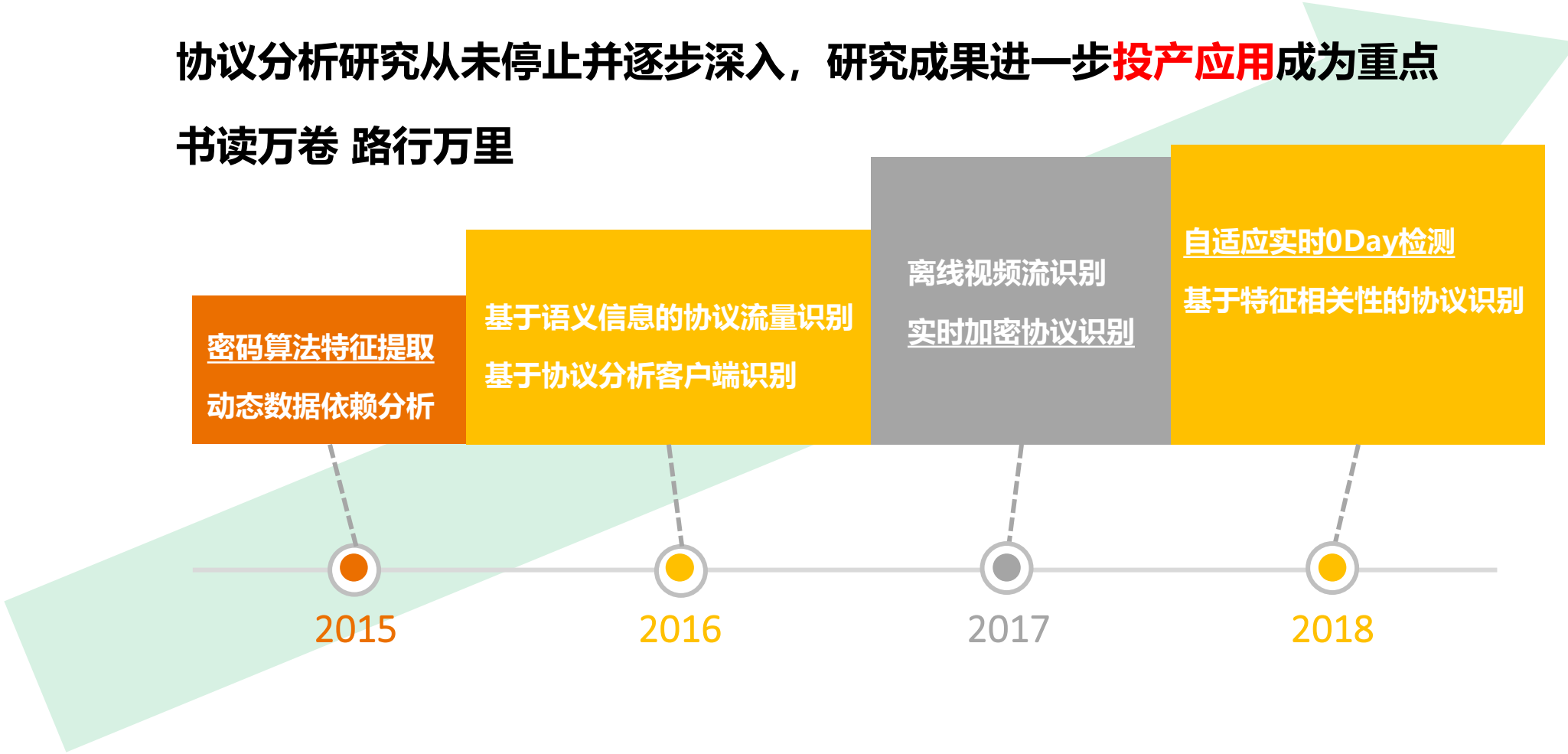
随难度系数攀升可以胜任的人越稀缺

高难度任务依然处于中常规战力无法解决的困境

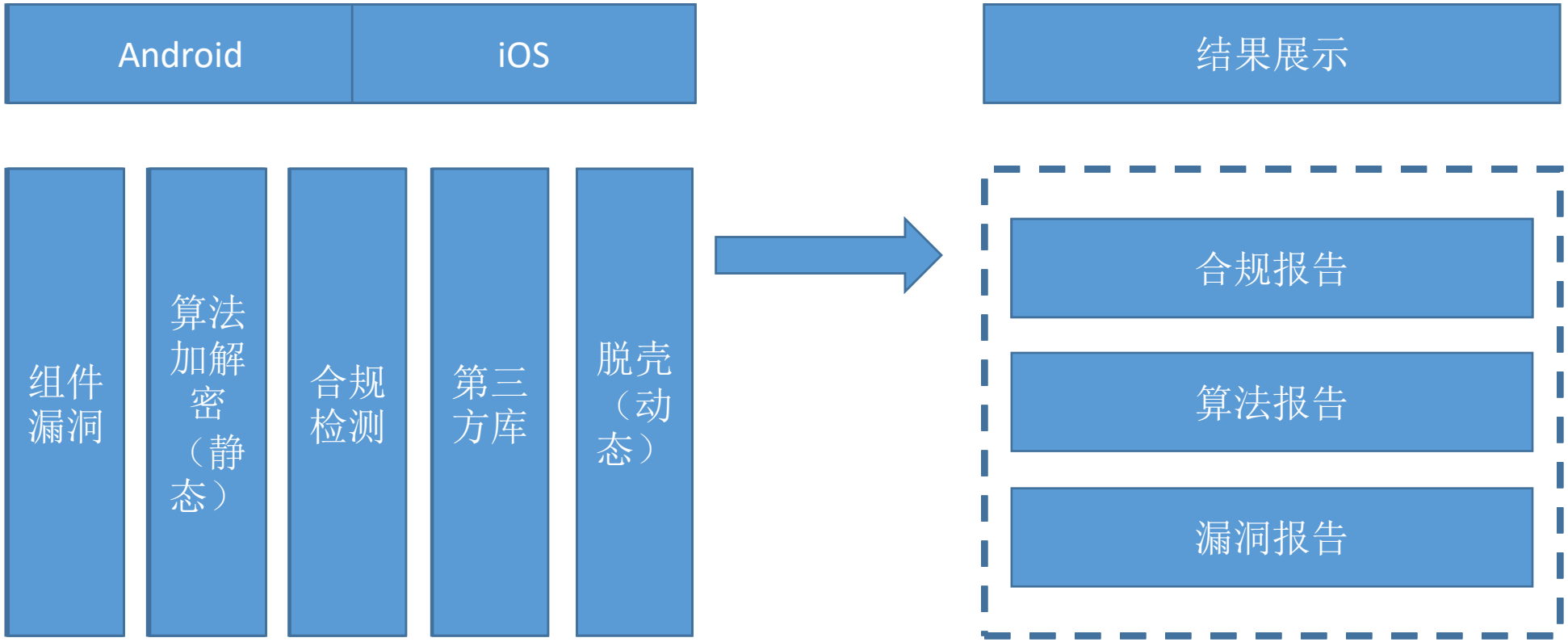
移动端对抗升级

协议分析研究从未停止并逐步深入，研究成果进一步投产应用成为重点

书读万卷 路行万里



应用场景（一）



应用场景（二）

极难静态分析

针对某一个功能点进行逆向分析，希望了解其实现原理；

有时会遇到这种情况：

- **代码严重混淆到极其复杂**
- **各种反静态分析手段**
- **返回值完全看不出什么**

```
000023EA ; =============== S U B R O U T I N E =======================================
000023EA
000023EA
000023EA                        public JNI_OnLoad
000023EA JNI_OnLoad             proc near
000023EA ; __unwind {
000023EA                        push    ebx
000023EB                        call    sub_2407
000023F0                        pop     ebx
000023F1                        add     ebx, 2Bh ;
000023F7                        push    ebx
000023F8                        retn
000023F8 JNI_OnLoad             endp ; sp-analysis
```

Decompilation failure:
1B7AE: positive sp value has been found

Please refer to the manual to find appropriate actions

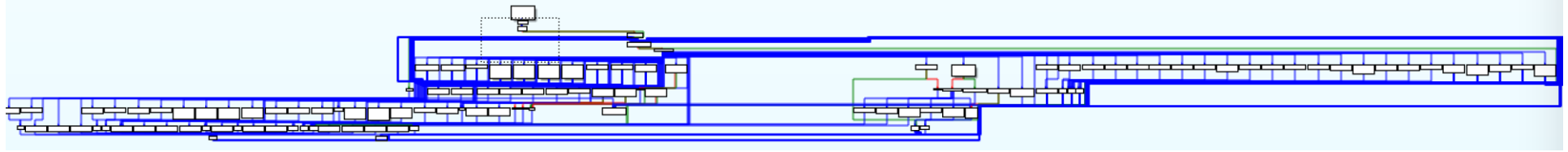
Don't display this message again (for this session only)

OK

```
static {
    System.loadLibrary(DIDLIB);
}

public static native String getDeviceID(long seed);
public static native String getDeviceData();
```

```
: [*] NCDid -> getDeviceID => AW009x1VIRtRm63k6LeC
: [*] NCDid -> getDeviceData => z9w+DETQza92Jj5O08AgM/...hKyP12blpTJ8FJbUgKunOG/ktWBgX2TIxUnMMDkJ+KYo4mUNSa2FxaSS+YkJoL2+R
```

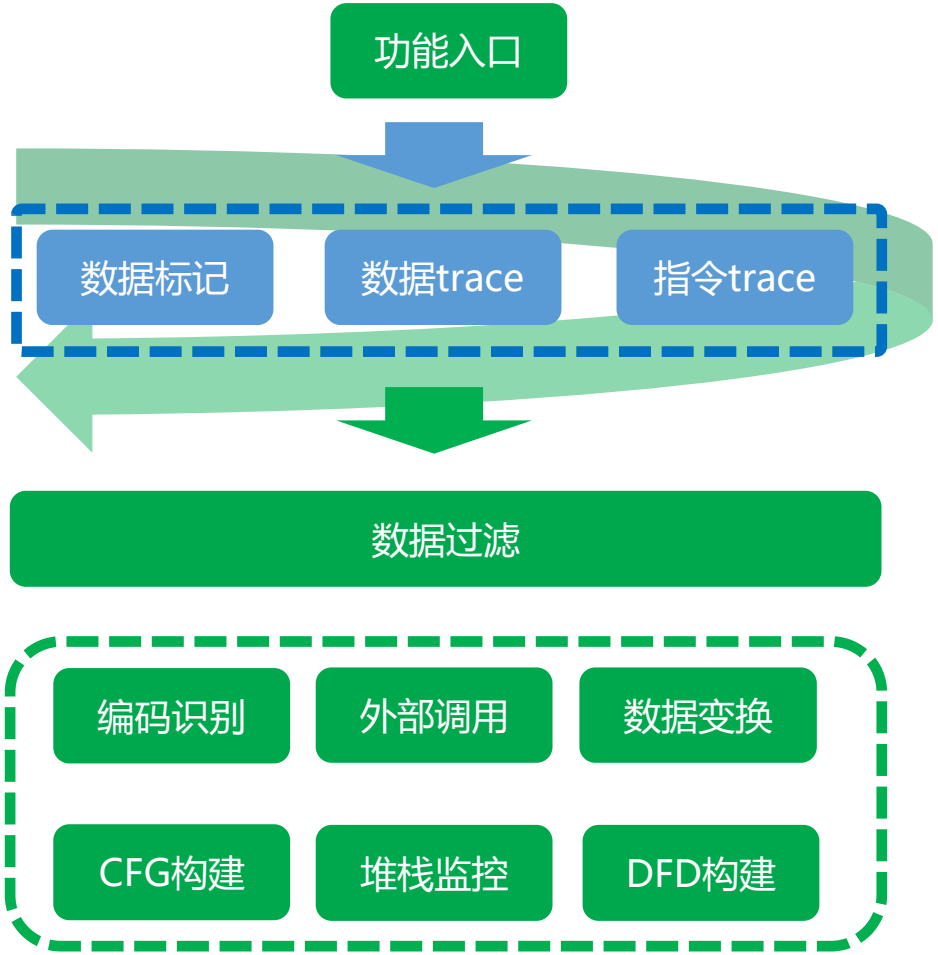


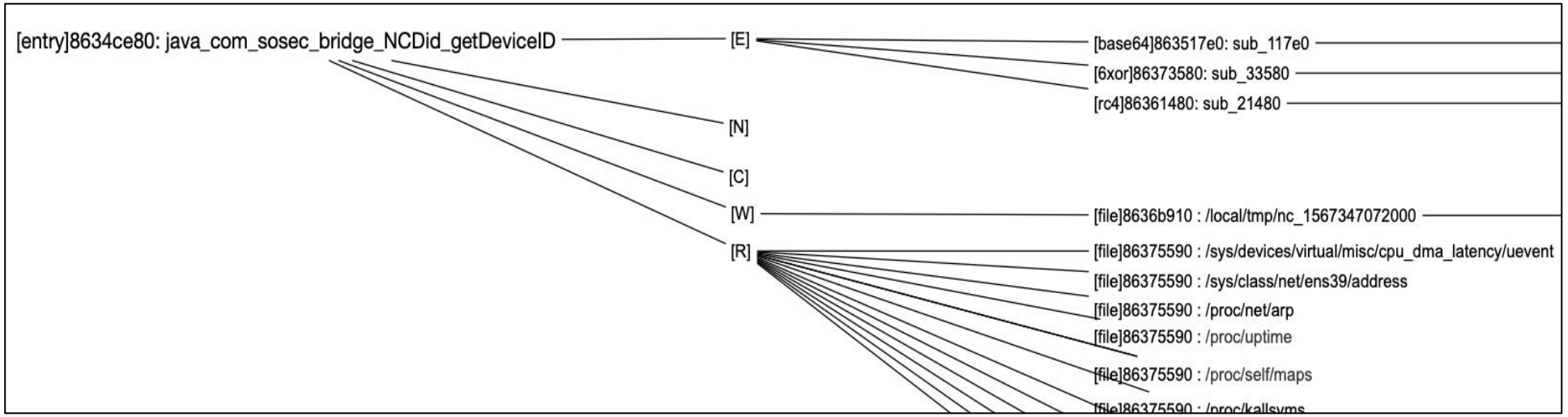
数据流分析

逆向分析中数据流分析十分适用于梳理某一功能实现；

通过对数据的自动追踪、运算匹配和指令trace，可为分析人员提供有价值的信息。

```
x14 = 0x0
x12 = 0x0
x13 = 0x1
x10 = 0x6426000f
x29 = 0x16e9ecfc0
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
lldb = 0x100743d60
ida = 0x100737d60
counter = 9540
pc = 0xcffff7d60
,0xb,0x1,0x0,0xb9     |disam_instruction(46):0x1000 :
hook_code(58):!!!!find str instruction!!!!: str
x8 = 0x16e9ecf20
x9 = 0x1c
x11 = 0x1
x14 = 0x1
x12 = 0x1
x13 = 0x0
x10 = 0x4
x29 = 0x16e9ecfc0
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
```





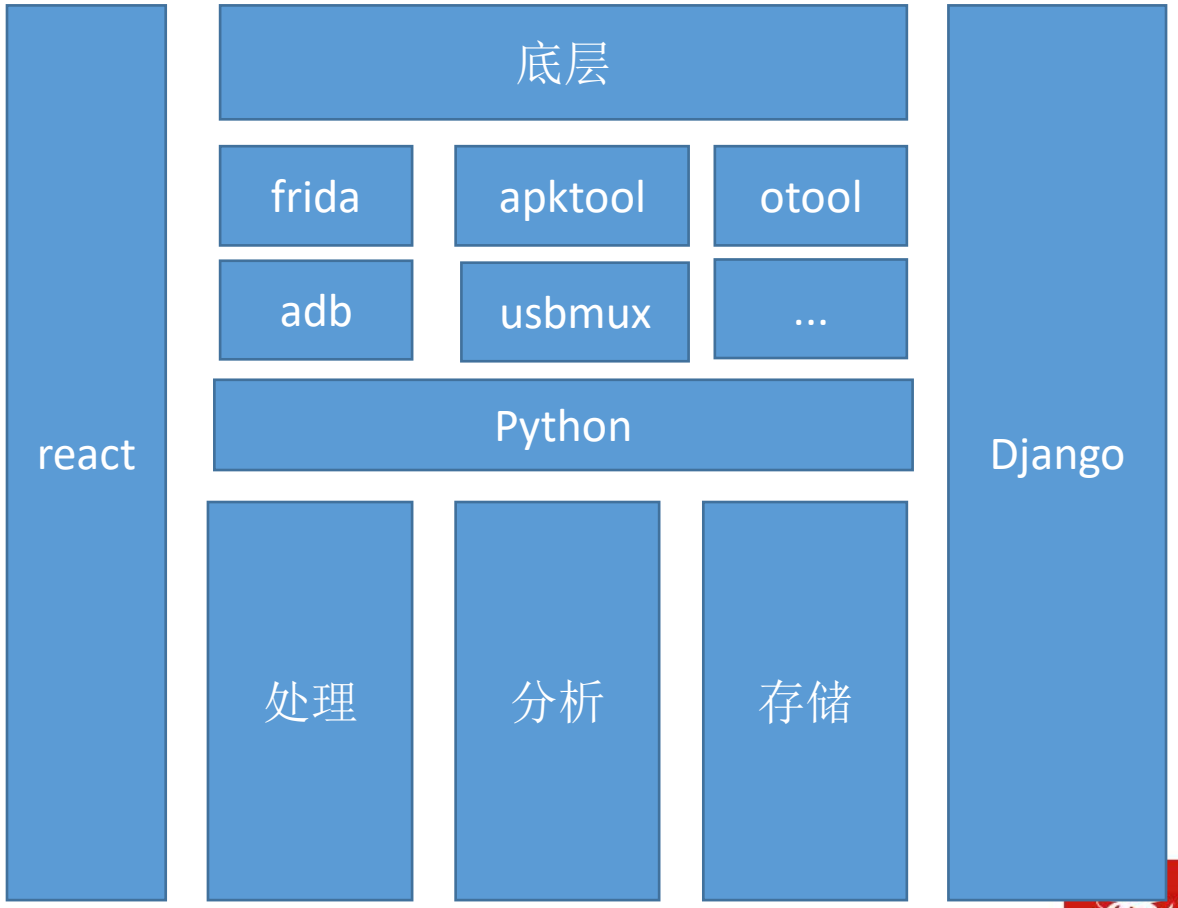
为分析人员进一步分析提供参考信息，包括：

文件读写、编码运算、外部调用以及网络等功能对应的符号地址和识别结果

为什么用到Python

轻量化、满足安服人员APP审计的需求、开发低成本

- 轻量化
- 满足基本安服需求
- 低成本
- 跨平台
- 易于移植



优化？效率？

前端	文本处理	数据存储方式	高并发
使用浏览器缓存，将一些常用的 css，js，logo 图标，这些静态资源缓存到本地浏览器， 通过设 置 http 头中的 cache-control 和 expires 的属性，可设定浏览器缓存，缓存时间可以自定义。 对 html，css，javascript 文件进行压缩，减少网络的通信量	尽量选择集合、字典数据类型，千万不要选择列表，列表的查询速度会超级慢，同样的，在已经使用集合或字典的情况下，不要再转化成列表进行操作 多使用iteritems()少使用items()，iteritems()返回迭代器	json是一种轻量级的数据交换格式。采用完全独立于编程语言的文本格式来存储和表示数据。层次结构简洁而清晰，易于人阅读和编写，同时也易于机器解析和生成，并有效地提升网络传输效率。	使用 nginx + uwsgi 提供高并发

总结与展望

总结：

随着互联网的发展，在移动安全领域逆向分析是不可或缺的一环，有着诸多的应用。如协议分析、算法识别、二进制对抗、加密流量分析。

自动审计平台赋能人工分析，大幅提高分析的效率、减少工作量，减少人员能力断层差。

展望：

推动二进制对抗进入新次元

1. 全自动化解决黑盒分析，常规代码混淆将会失效，将会有新形式的代码混淆保护方案；

THANKS!

感谢观看