RSA CONFERENCE
CHINA 2012
RSA信息安全大会2012

# THE GREAT CIPHER

## MIGHTIER THAN THE SWORD

## 伟大的密码胜于利剑

# 移动互联网时代，应用识别技术面临的新挑战和解决方案展望

演讲人姓名：严雷

演讲人公司：网康科技

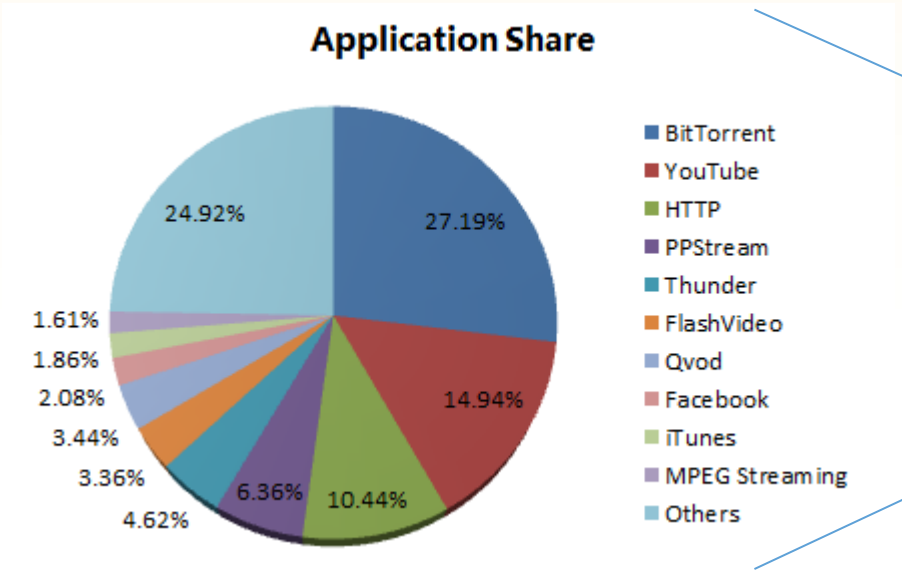
RSA CONFERENCE CHINA 2012

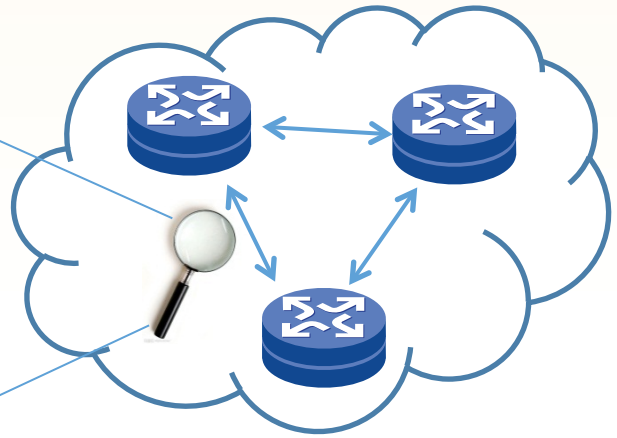
RSA信息安全大会2012

# 应用识别



应用识别 -> 识别哪些应用”跑在”网络中



为什么需要应用识别?

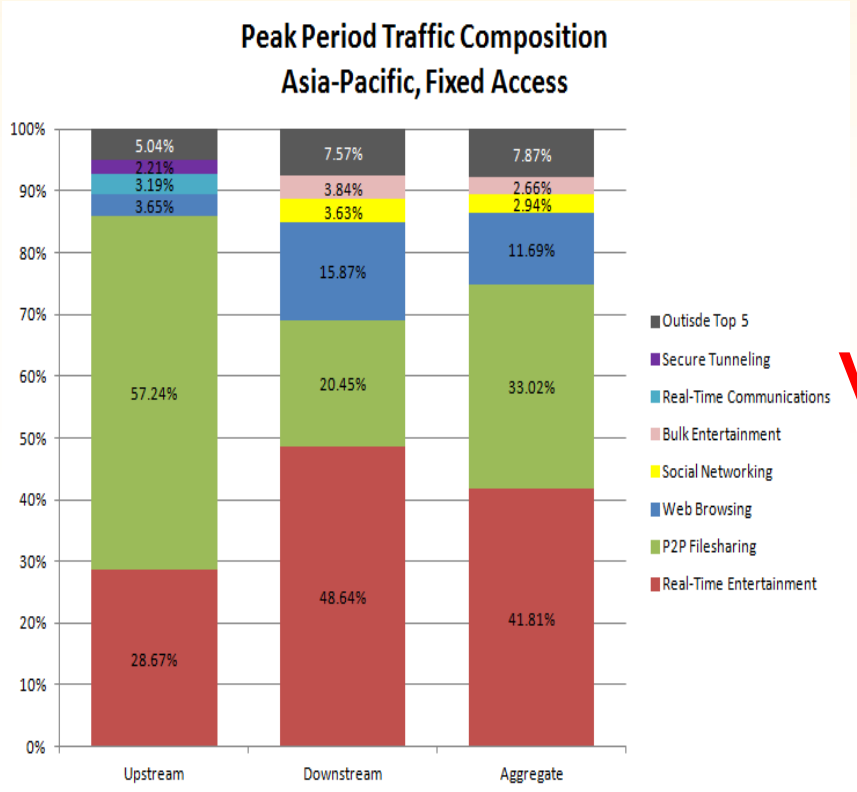
- 网络管理 – 带宽管理，网络规划，计费等
- 网络安全 – 下一代防火墙，IDS/IPS等

# 应用识别技术回顾

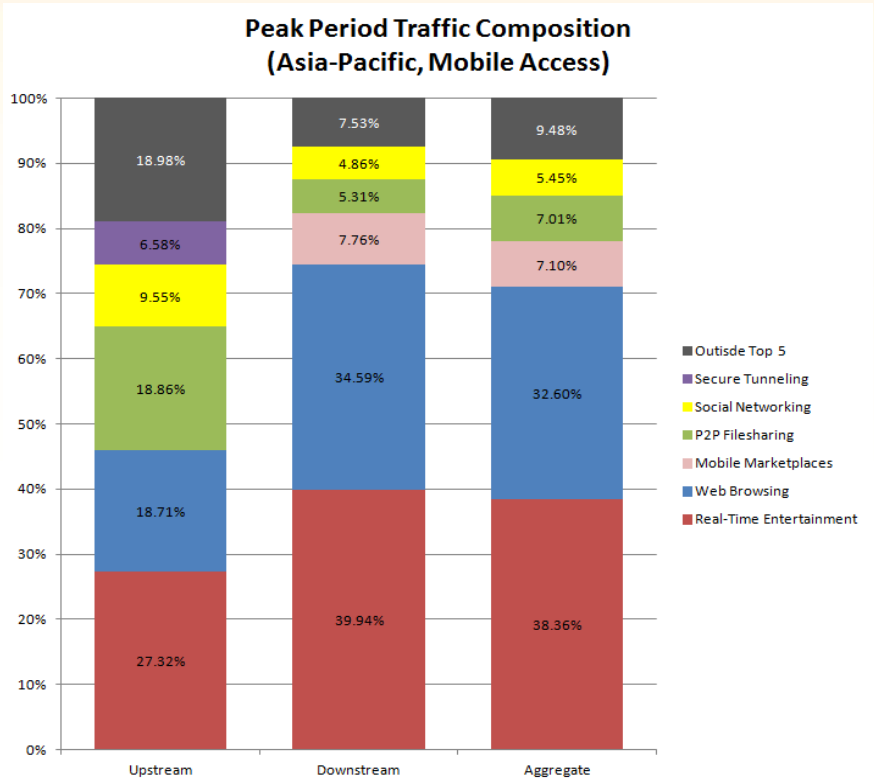
RSA CONFERENCE CHINA 2012

| 基于端口特征识别 | DFI (基于流特征识别) | DPI (基于内容识别) | 定制化识别 |
| --- | --- | --- | --- |
| • 典型应用: E-mail, DNS, FTP, HTTP<br>• 优点: 性能高<br>• 缺点: 应用端口复用普遍， 识别准确率低 | • 典型应用: P2P (Bitorrent)<br>• 优点：性能高， 能对加密应用分类<br>• 缺点：准确率低 | • 典型应用：最重要应用识别方法， 适用各种应用种类<br>• 优点：准确率高<br>• 缺点：性能低， 人工生产特征 | • 典型应用：Skype, Xunlei,<br>• 优点：准确率高<br>• 缺点：方法适用性窄, 通常要Hard-Code |

# 移动互联网应用流量分布



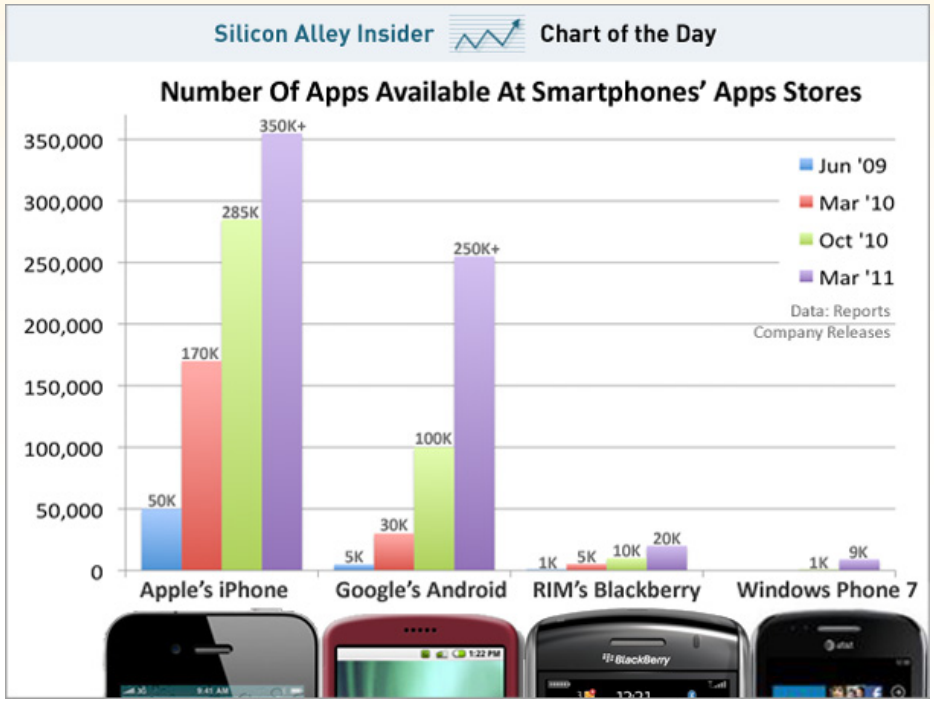

VS




2012年亚太地区固网与移动网络流量分布比较

# 变化一: 移动互联网应用爆炸式增长

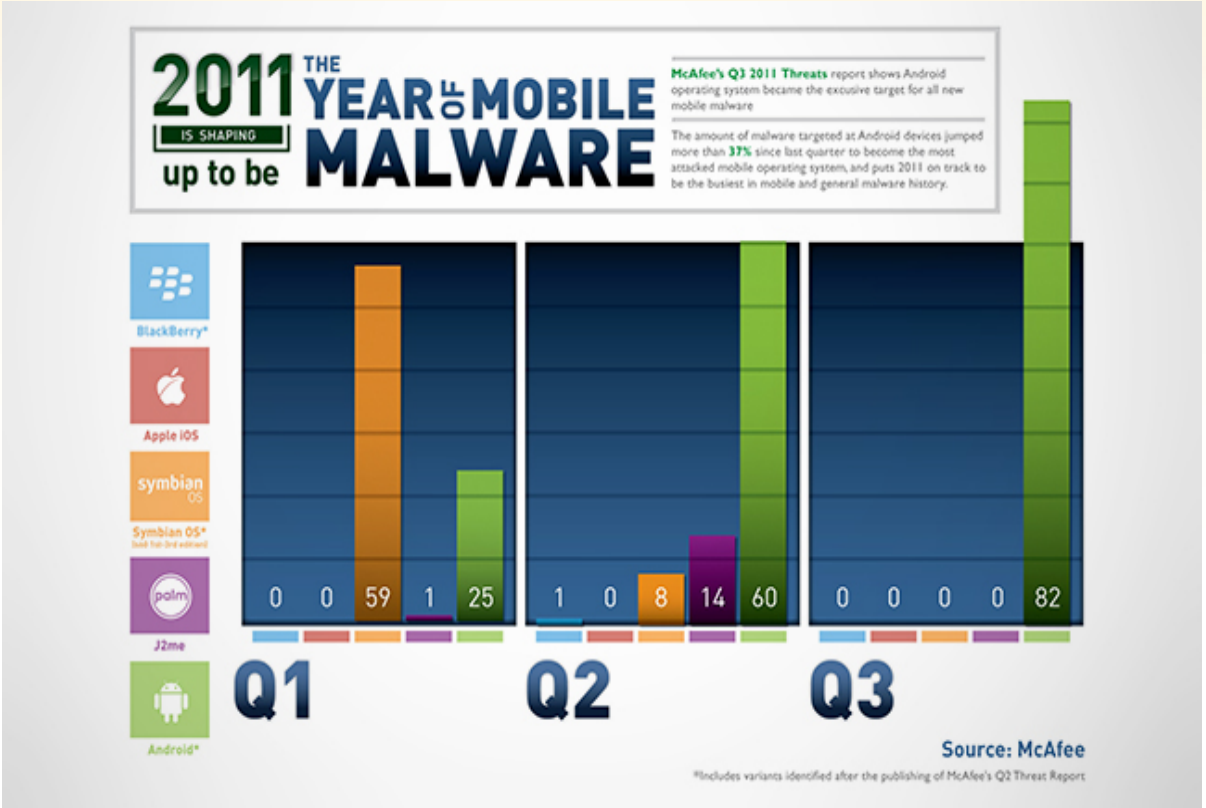
RSA CONFERENCE CHINA 2012




- Apple AppStore 500K+ 应用，并且平均每月新增应用数 2.3万
- Android Market 400K+ 应用，并且平均每月新增应用数～2万
- 但目前应用特征库提供厂商，一般更新频率在2周～1个月，每次更新大概几个到几十个应用特征，大大落后于每月新增应用数。

# 变化二: 移动设备成为攻击者新的战场

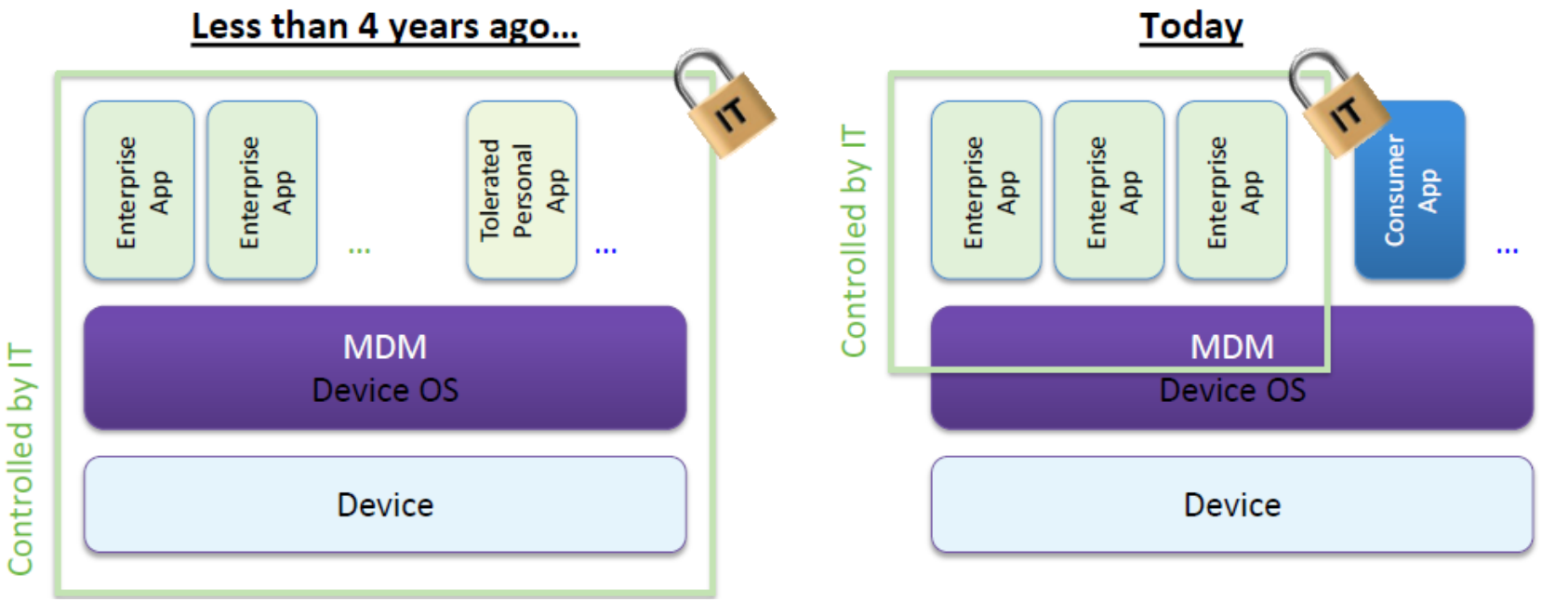





- 2011年24,794恶意软件被发现，相对2010年增加367%（6760），相对2009增加503%（1649）
- 2001年，约1800万台Android被感染，其中中国占31.6%.

# 变化三：企业需要对运行在网络中的移动应用进行控制

RSA CONFERENCE CHINA 2012

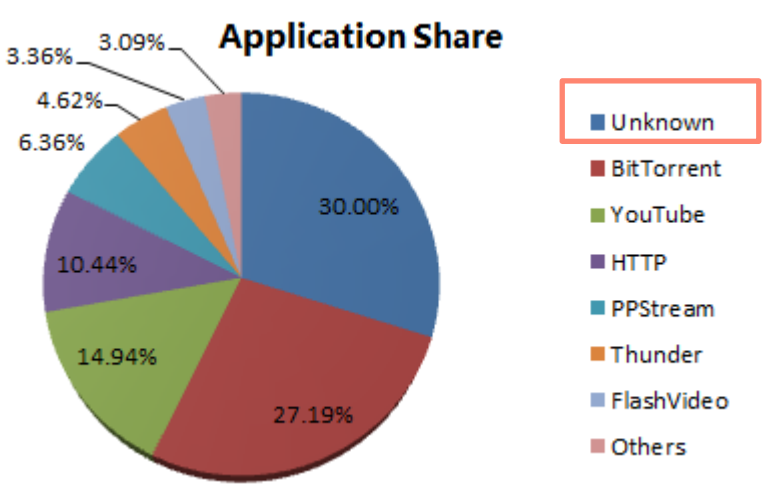



- 支持BYOD(允许带自己移动设备办公), 正成为企业IT不得不面对解决的问题
  - 企业内，只有50%的智能手机和20%的tablets是由IT购买的. （IDC Survey, 2010)
- 企业移动设备管理市场规模将达到12.B$ by 2015 (ABI Research)

# 挑战一：0-day 应用

RSA CONFERENCE CHINA 2012

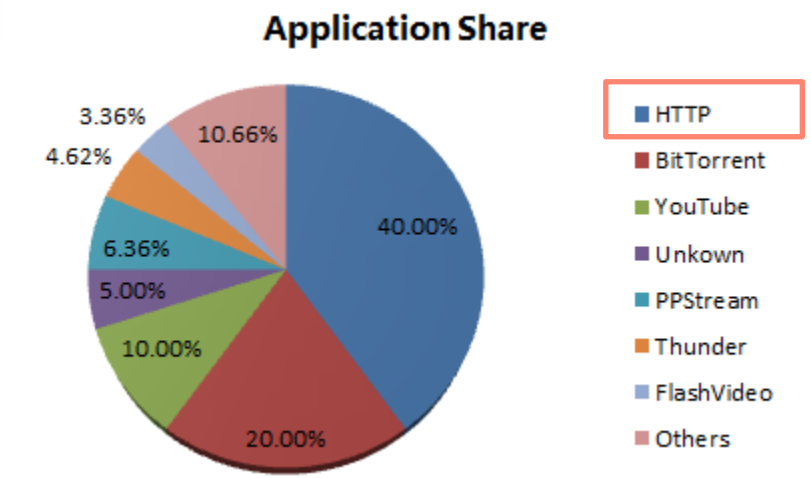
- 以前
  - 需要支持几百个应用
  - 应用更新大概2周～1个月
- 现在
  - 现在需要支持几千个应用，甚至更多
  - 非常快应用更新频率，比如：~1天，甚至实时
- 挑战
  - 协议库生产效率
  - DPI技术失效



# 挑战二：应用细分



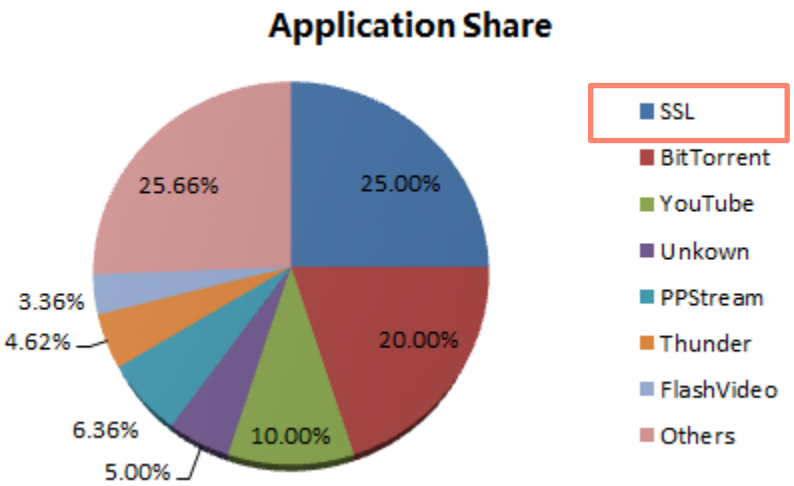
- 以前：
  - 大多数C/S应用
  - 应用数目Top 50~100
- 现在
  - 绝大部分应用基于HTTP协议， 应用数目巨大
  - 几个平台应用上， 又承载了大量的应用, 比如: QQ包括: QQ聊天,QQ文件传输, QQ游戏等;
- 挑战
  - 大量Http应用的细分， 传统URL分类方法在Mobile平台上失效
  - 平台应用的细分， 连接复用/关联等特点增大复杂度



# 挑战三：规避检测技术



- 以前：
  - 大多数是明文协议
  - 应用开发时，不会考虑规避检测
- 现在：
  - 加密
  - 检测是否被控制，动态调整流特征
- 挑战
  - 加密 -> DPI技术失效
  - 隐藏Flow特征 -> DFI技术失效，比如：包长检测技术



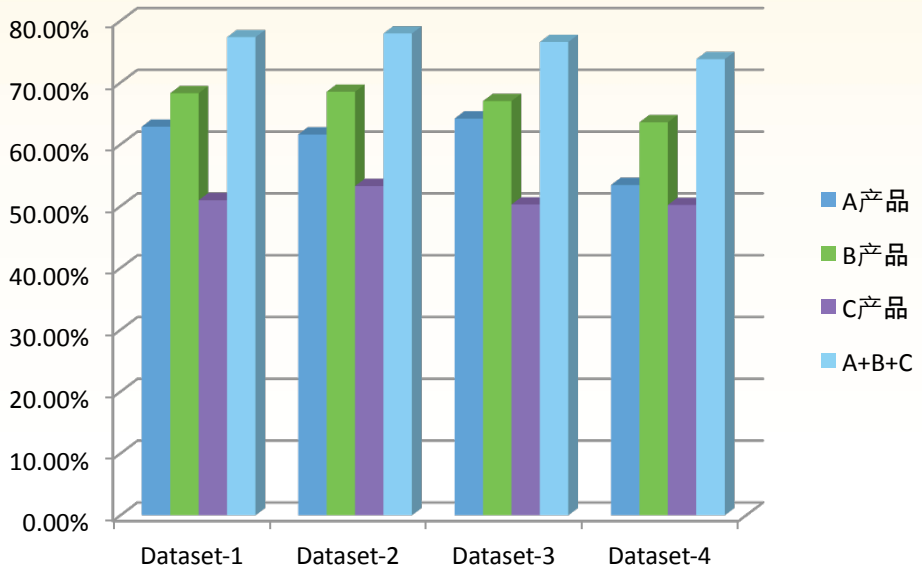
# 解决方案一: 应用识别自动化




- 自动化特征提取生产过程
  - 应用的下载安装
  - 应用流量产生
  - 自动特征提取
- 自动特征提取技术
  - 协议解析方向工程
  - 特征模式: 大多数Http应用会用user-agent标识等;

# 解决方案二: 使用多应用识别引擎

RSA CONFERENCE CHINA 2012




| | DFI特征数 | DPI特征数 | 定制化识别 |
|---|---|---|---|
| A产品 | 低 | 500~ Apps | 低 |
| B产品 | 高 | 300~ Apps | 低 |
| C产品 | 低 | 700~ Apps | 低 |

多识别引擎可以显著提高识别率和准确率.

# 解决方案三：扩展DFI技术



| 用户样本 | 应用数量 | DPI 识别率 | 聚类后识别率 | 聚类准确度 |
|---|---|---|---|---|
| 1 | 7 | 93.60% | 97.32% | 98.20% |
| 2 | 10 | 75.40% | 91.37% | 97.90% |
| 3 | 8 | 72.20% | 96.75% | 98.00% |
| 4 | 10 | 90.50% | 98.21% | 97.90% |

某网络，User-Local聚类识别效果

- 探索流间(连接间) 特征，代替使用连接内数据包间的特征，比如：
  - 服务Cache（目的IP+Port)
  - 连接间使用端口Sequence
- 改进聚类和机器识别的算法
  - 比如: 使用User-Local 聚类 代替 全局聚类，降低无识别率和范围.

# 总结



- 近年来，移动互联网的快速发展，对应用识别带来新的挑战
  - 0-day 应用
  - 应用细粒度
  - 规避检测
- 面对挑战，应用识别技术可以从以下几个方向提供新的解决方案
  - 自动化应用识别技术
  - 多应用识别引擎
  - 扩展DFI技术

# 谢谢

RSA CONFERENCE
CHINA 2012
RSA信息安全大会2012