2016 中国互联网安全大会
China Internet Security Conference

协同联动 共建安全+命运共同体

大数据视角下的Web威胁分析

林明峰 安恒信息北方区技术总监

WEB在线威胁领域

外网 会话开始 登录 交易 退出

Web 威胁领域

- 钓鱼网站
- 网站盗取
- 漏洞探测
- 第七层DDoS攻击

- 密码猜测、盗用
- 参数注入
- 新账号注册欺诈
- 高级恶意软件 (木马)
- 促销滥用

- 中间人、浏览器中间人
- 账号接管
- 未授权账号交易
- 非法资金转移

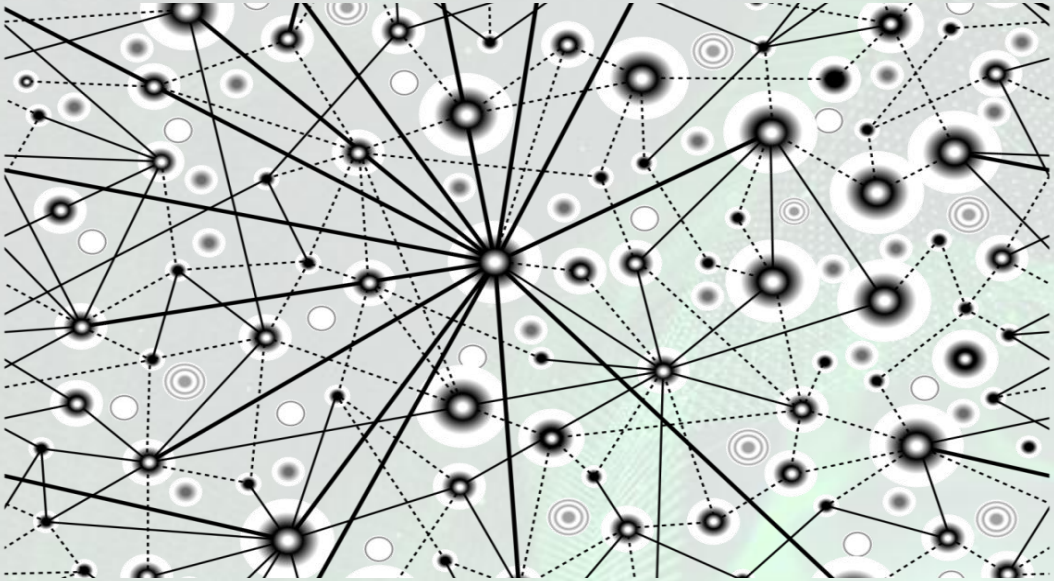
信息安全

欺诈

身份验证前的威胁

身份验证后的威胁

我们今天怎么保护网站应用？



用户

- 用户名、密码
- 双因子认证
- 设备认证

网络

- 防火墙
- IPS/IDS

应用

- WAF
- 渗透测试
- 动态扫描
- 日志分析、SIEM
- 源代码分析

对用户行为没有可视性！

用户在网站上做什么？

下单!

我只是随便看看…

我有这么多账号，一个一个试一下

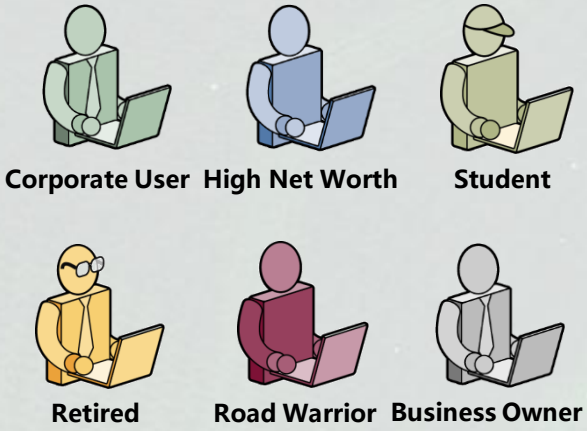
手机、平板上都能买卖了，都不用登录

我需要转一笔钱

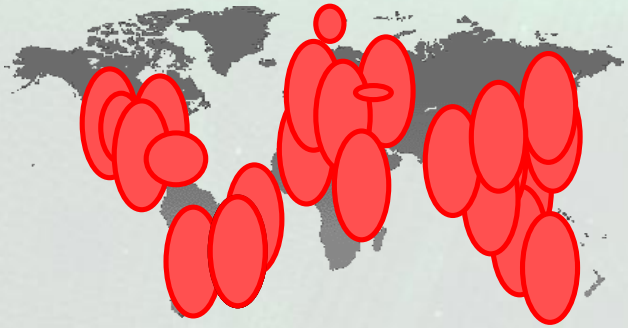
正常交易？破坏性的？非法的？

Web威胁检测服务系统的概念

用户总是有各种类型



… 黑客的位置飘忽不定



以用户为核心的身份认证和欺诈检测

- 对大多数用户是不可见的
- 非侵入式 (不需要安装任何软件在客户端)
- 了解每一个用户独特性
- 学习用户的行为习惯

智能和异常检测

- 学习和适应每一个用户独特行为
- 强制性策略阻止攻击人和欺诈者
- 第一时间阻止新型攻击
- 针对在线攻击独特的视角

什么是Web威胁检测系统

- 它是一个利用革命性的网络安全 Web 会话情报进行实时威胁检测的系统；透过先进的可视化数据分解，提供有意义的结论，并且根据这些结论去采取行动

- 采取威胁指数分析引擎技术
  - 针对用户的整个访问周期内（登入前、中、后）所有的行为进行威胁评估
  - 根据威胁指数、策略、或群体或个人行为模式生成时间与警报

实时连接用户行为与技术分析

基于异常行为分析
为群体和个人建立动态行为模型

Web 会话可视性
让 “噪音” 成为可执行的结论

流动式分析
利用Spark强壮大数据框架可以实时提供可视化的智能分析和基于风险的行为威胁检测

能够将整理出的会话数据及分析以流线方式传输到外部数据湖

跨渠道
在不同设备和渠道里对用户执行连续性的监控

威胁评估引擎
速度、中间人、浏览器中间人、行为、时间顺序等指数

在线防欺诈系统如何进行工作

电子交易系统

Low/High

Allow Challenge

基于风险威胁检测引擎

规则引擎 行为引擎

登入数据

评分

风险评价

工单管理

认证挑战

检验

通过/失败

介入检测

介入检测的手段

它是如何工作的？


登入前行为
登入中行为
登入后行为
Web 威胁分析引擎
分析引擎(0-100)
中间人 20
浏览器中间人 60
速度 100
参数 0
行为 90
匿名 10
规则引擎
EDS
警报和事件管理
行动服务器
邮件
案件管理
日志 SIEM
负载均衡器
WAF
API 操作
整个会话数据点击流分析
管理界面
仪表盘
历史行为的时间表
搜索及报告
数据流（外置数据平台）

客户端设备指纹与业务数据结合

设备指纹

交易数据

网络信息

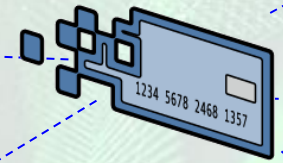
IP Address
ISP
Connection Type
Is Mobile Gateway

地理位置

Country
City

设备信息

OS & OS Version
Language
Cookie
Browser info
Device ID



在线交易

用户信息

User ID (HASH)
User Category

交易信息

Transaction Type
Server Time
Amount

时间信息

Month, Week
Day, Hour
Frequency

任何参数的组合
都可以使用

设备指纹在防欺诈中应用

用户身份鉴别

设备唯一性 | 环境有效性

↓

用户行为建模

识别用户场景 | 区别例外情况

↓

实时规则过滤

已知异常场景 | 业务规则

↓

实时风险分析

场景异常 | 新场景生成

设备指纹的信息分类

硬件信息	软件信息	网络信息	位置信息
硬件ID（HardwareID）	语言（Languages）	无线MAC地址（WiFiMacAddress）	地理位置：海拔高度（GeoLocation: Altitude）
手机号码（PhoneNumber）	时间戳（TIMESTAMP）	CellTowerId	地理位置：海拔高度准确度（GeoLocation: Altitude Accuracy）
屏幕尺寸（ScreenSize）	多任务支持（MultitaskingSupported）	本地区域代码（LocationAreaCode）	地理位置：前进方向（GeoLocation: Heading）
SIM_ID	操作系统ID（OS_ID）	移动国家代码（MCC）	地理位置：水平精确度（GeoLocation: HorizontalAccuracy）
设备模式（DeviceModel）	应用程序Key（ApplicationKey）	移动网码（MNC）	地理位置：纬度（GeoLocation: Latitude）
设备名称（DeviceName）	SDK版本（SDK Version）	无线网络数据：基本服务标识（WiFiNetworksData: BBSID）	地理位置：经度（GeoLocation: Longitude）
设备系统名称（DeviceSystemName）		无线网络数据：通道（WiFiNetworksData: Channel）	地理位置：速度（GeoLocation: Speed）
设备系统版本（DeviceSystemVersion）		无线网络数据：信号强度（WiFiNetworksData: SignalStrength）	地理位置：状态（GeoLocation: Status）
		无线网络数据：服务标识（WiFiNetworksData: SSID）	地理位置：时间戳（GeoLocation: Timestamp）
		无线网络数据：站点名称（WiFiNetworksData: StationName）	

设备唯一性

交易环境有效性

设备指纹的唯一性分析

- 基本不发生变化
- 考虑哈希处理用户敏感信息
- 协助确认唯一用户身份
- 形成唯一化的用户标识

硬件信息	软件信息
硬件ID（HardwareID）	语言（Languages）
手机号码（PhoneNumber）	时间戳（TIMESTAMP）
屏幕尺寸（ScreenSize）	多任务支持（MultitaskingSupported）
SIM_ID	操作系统ID（OS_ID）
设备模式（DeviceModel）	应用程序Key（ApplicationKey）
设备名称（DeviceName）	SDK版本（SDK Version）
设备系统名称（DeviceSystemName）	
设备系统版本（DeviceSystemVersion）	

UID（用户唯一标识）

- 可能发生变化
- 无需哈希处理
- 设备指纹辅助信息

设备指纹在防欺诈中应用

用户身份鉴别

设备唯一性 | 环境有效性

用户行为建模

识别用户场景 | 区别例外情况

实时规则过滤

已知异常场景 | 业务规则

实时风险分析

场景异常 | 新场景生成

用户行为建模方式

输入数据源

- 用户唯一标识（UID）
- 软件信息
- 网络和位置信息
- 交易信息

建模算法

- 聚类算法
- Rock 算法简化版本

输出方式

- 形成用户行为模型
- 对例外情况需要进行修正

聚类分析（Cluster Analysis）

ISC 2016 中国互联网安全大会 360互联网安全中心

聚类分析（Cluster Analysis）又称群分析

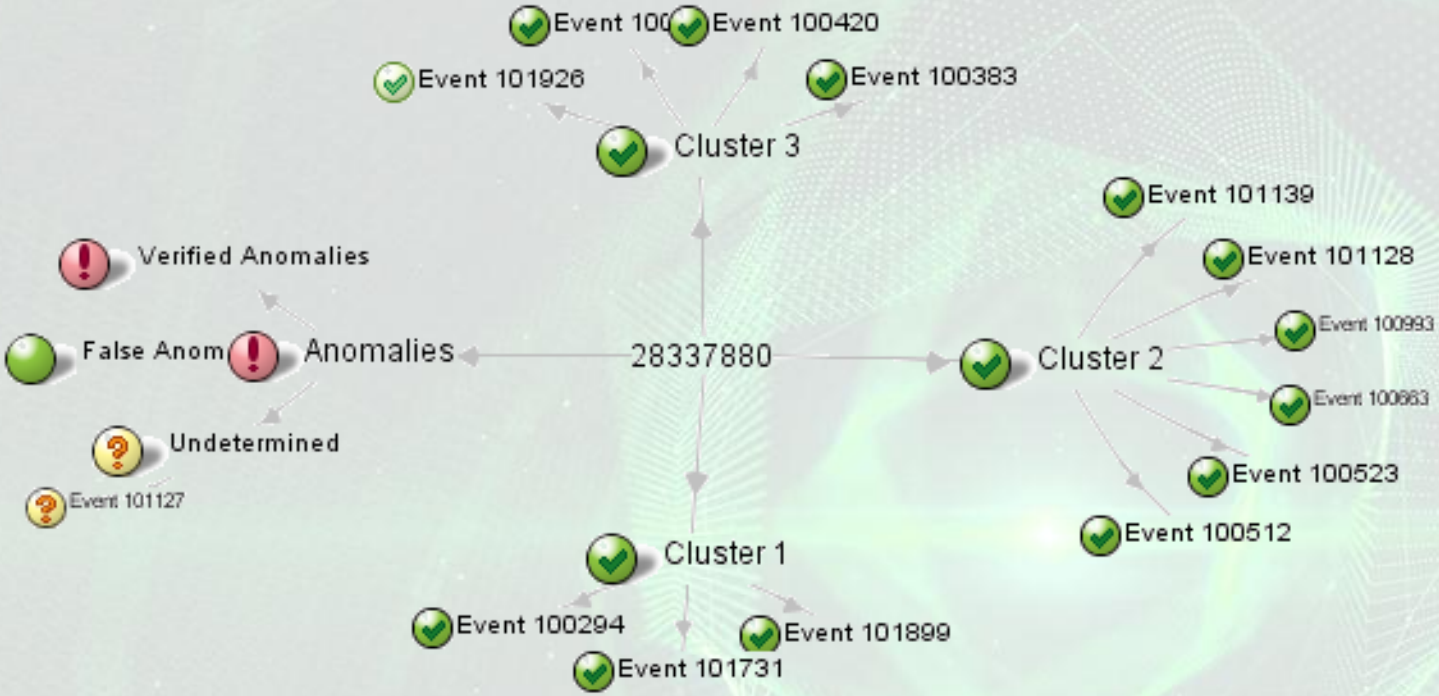
- 根据“物以类聚”的道理，对样品或指标进行分类的一种多元统计分析方法，它们讨论的对象是大量的样品，要求能合理地按各自的特性来进行合理的分类，没有任何模式可供参考或依循，即是在没有先验知识的情况下进行的。
- 随着人类科学技术的发展，对分类的要求越来越高，以致有时仅凭经验和专业知识难以确切地进行分类，于是人们逐渐地把数学工具引用到了分类学中，形成了数值分类学，之后又将多元分析的技术引入到数值分类学形成了聚类分析。

聚类是将数据分类到不同的类或者簇这样的一个过程

- 同一个簇中的对象有很大的相似性，而不同簇间的对象有很大的相异性。聚类分析的目标就是在相似的基础上收集数据来分类。
- 聚类源于很多领域，包括数学，计算机科学，统计学，生物学和经济学。在不同的应用领域，很多聚类技术都得到了发展，这些技术方法被用作描述数据，衡量不同数据源间的相似性，以及把数据源分类到不同的簇中。

用户行为分析图

聚类算法产生的用户行为模式



设备指纹在防欺诈中应用

用户身份鉴别

设备唯一性 | 环境有效性

↓

用户行为建模

识别用户场景 | 区别例外情况

↓

实时规则过滤

已知异常场景 | 业务规则

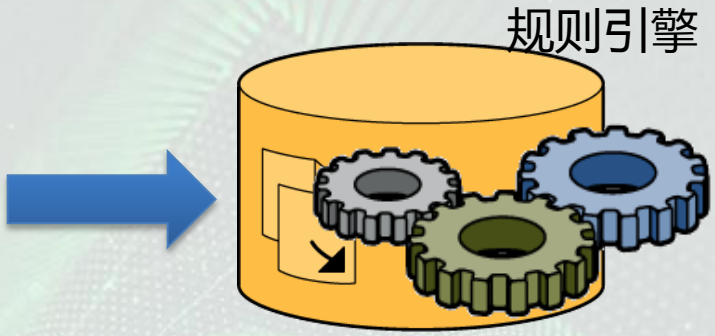
↓

实时风险分析

场景异常 | 新场景生成

针对已经异常场景分析

网络信息	位置信息
无线MAC地址（WiFiMacAddress）	地理位置：海拔高度（GeoLocation：Altitude）
CellTowerId	地理位置：海拔高度准确度（GeoLocation：Altitude Accuracy）
本地区域代码（LocationAreaCode）	地理位置：前进方向（GeoLocation：Heading）
移动国家代码（MCC）	地理位置：水平精确度（GeoLocation：HorizontalAccuracy）
移动网码（MNC）	地理位置：纬度（GeoLocation：Latitude）
无线网络数据：基本服务标识（WiFiNetworksData：BBSID）	地理位置：经度（GeoLocation：Longitude）
无线网络数据：通道（WiFiNetworksData：Channel）	地理位置：速度（GeoLocation：Speed）
无线网络数据：信号强度（WiFiNetworksData：SignalStrength）	地理位置：状态（GeoLocation：Status）
无线网络数据：服务标识（WiFiNetworksData：SSID）	地理位置：时间戳（GeoLocation：Timestamp）
无线网络数据：站点名称（WiFiNetworksData：StationName）	



数据输入

- 网络信息
- 位置信息

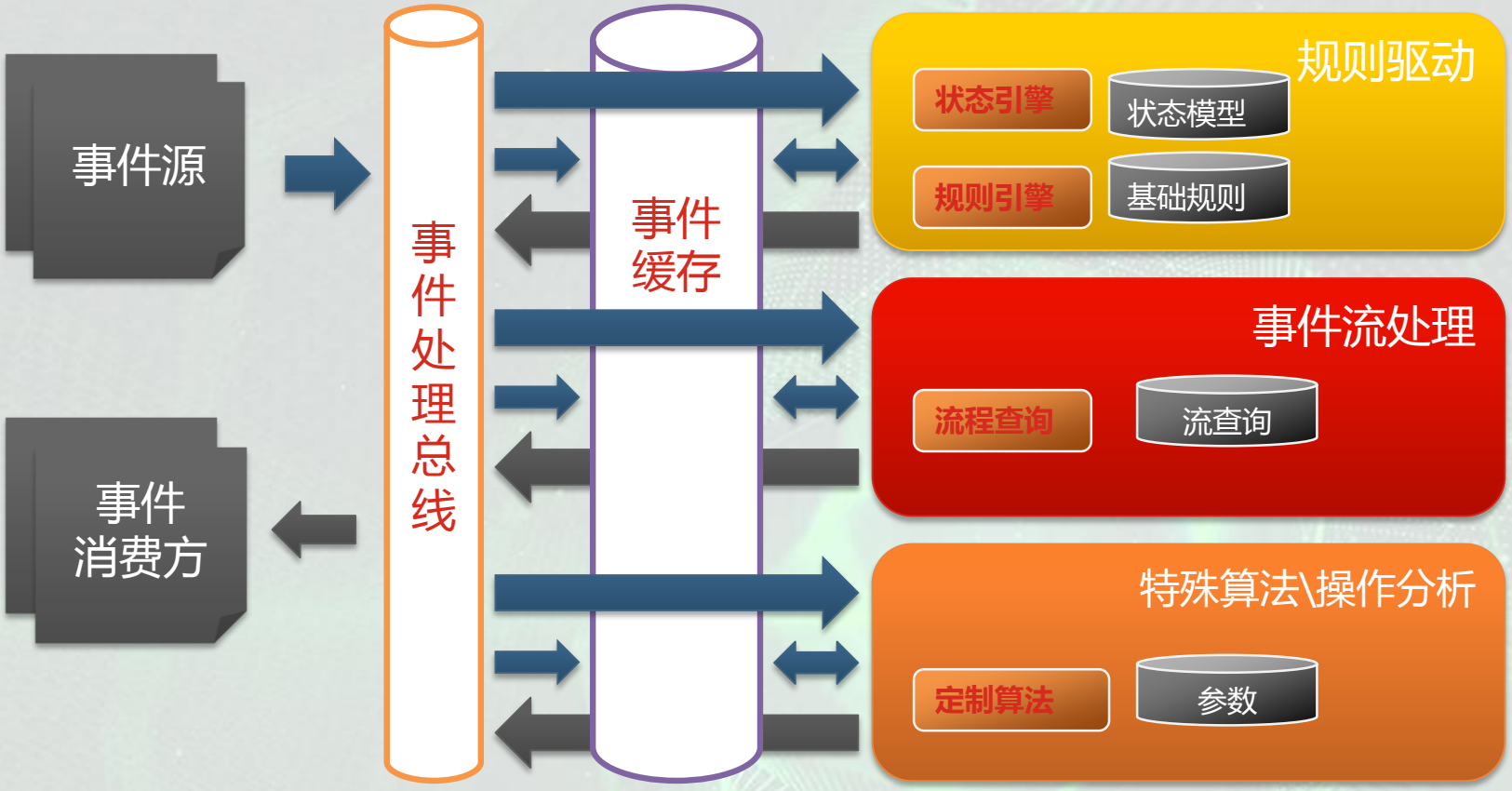
规则引擎

- CEP（complex event processing ）引擎

输出结果

- 已知业务违规
- 违规终端信息

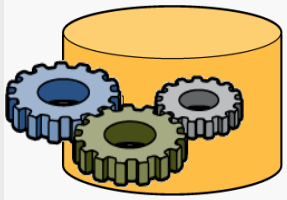
设备指纹和交易在CEP中的位置



大量的实时数据作为事件源接入**事件处理总线**，CEP引擎通过指定的规则，处理这些实时数据和缓存的历史数据，并通过事件处理总线将有意义的事件提供给事件消费方。

业务异常规则

Policy 零售网银规则1



IF (IP Address is Blacklisted) THEN Fraud Detected

IF (Logins From 2+ Locations are >1000 Miles Apart and Within 5 Minutes) THEN Fraud Detected

Additional Fraud Rules

cklisted IP Addresses

15.176.125.251
196.200.56.111
25.216.211.108
106.250.11.213
25.216.211.108
106.250.11.213
78.102.221.154
154.251.26.143
18.174.233.144
214.222.19.196
25.209.243.233
221.174.23

Verify the Distance Between Log

登录位置信息 1000 Miles

设备指纹在防欺诈中应用

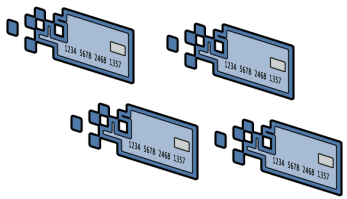
自学习的客户端指纹识别检测引擎

认可行为

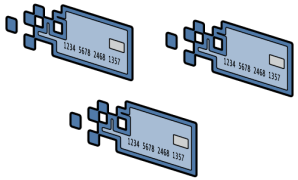
ANOMALY !

异常行为

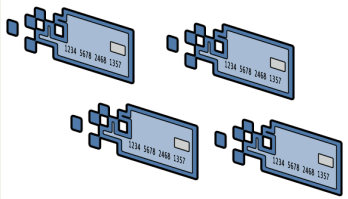
用户 JohnDoe01 行为模式(自动生成通过FDS)



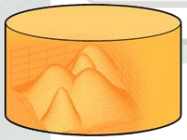
Transactions From Home After 8PM - Weekday



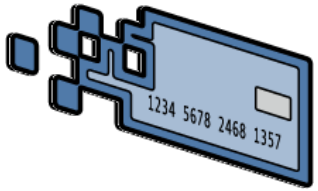
Transactions From Mobile Between 8AM and 5PM - Weekday



Transactions From Home on Weekend



JohnDoe01 Transaction2



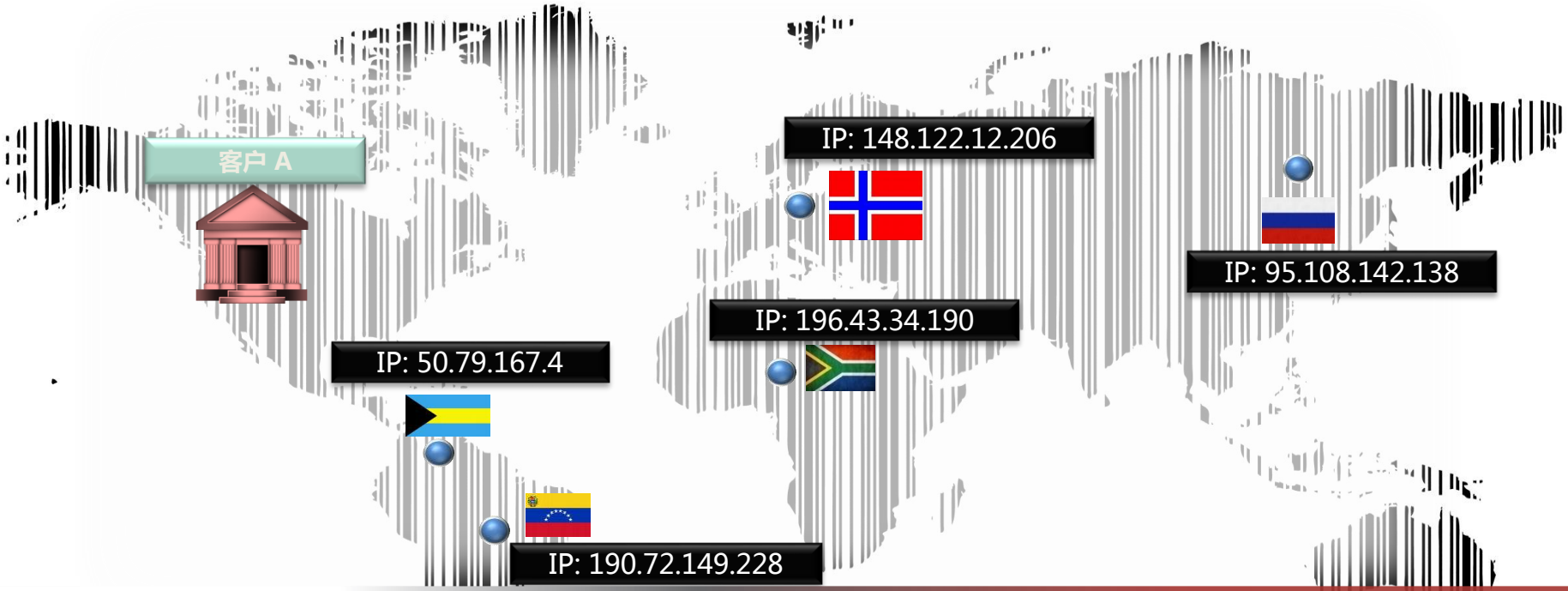
Transaction From Mobile Between 8AM and 5PM on Mon On Sun

介入控制和检验

- 策略和风险驱动介入方式
  - 规则被触发
  - 行为检测引擎返回高风险和高确认
- 介入方式有多种选择
  - 认证确认方式
    - OTP Credential
    - Out-of-band using VIP Service
  - 安全的问题与回答
  - 个性化的图片与回答
  - 客户中心呼叫或者管理员提醒
- 介入的策略是能被定制的
  - 风险适中，但是高价值的用户，需要选择介入
  - 呼叫中心对高端客户应该呼叫提醒他们当遇到高风险的时候

在线欺诈检测针对移动客户

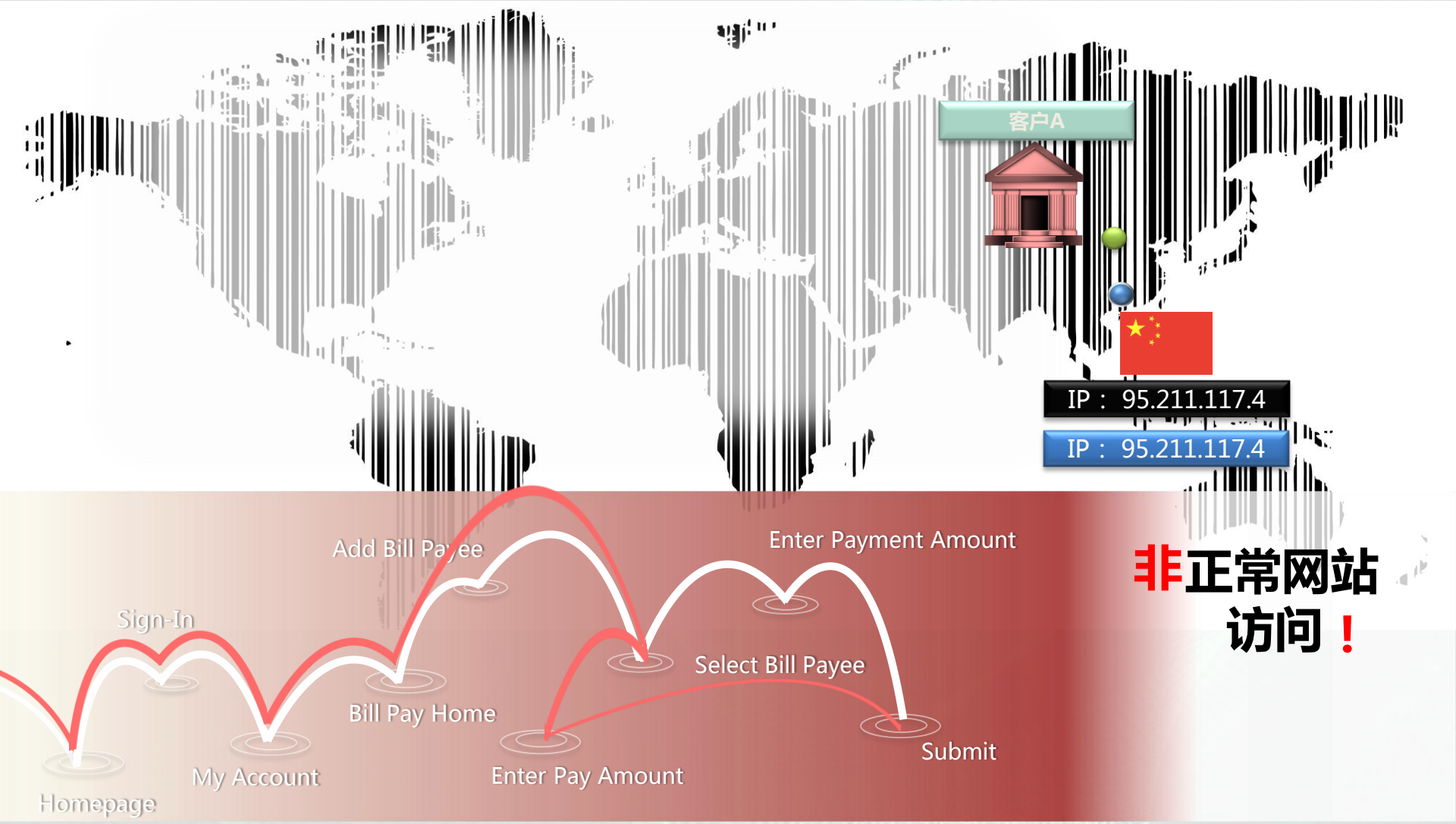
- 无需单独的实例或者通道
  - User ID, geolocation, connection type, IP address, browser type, OS 等信息是在移动设备和计算机设备中共享的信息
- 灵活的定制针对移动客户的个案
  - Connection type, isMobileGateway
- 支持设备ID或者定制化的移动用户参数
  - 无需对JavaScript的脚本形成依赖
  - 嵌入手机的身份认证程序将能提供将能提供动态的设备指纹
- 用户行为引擎将会更新用户特征，使用移动设备的模式识别



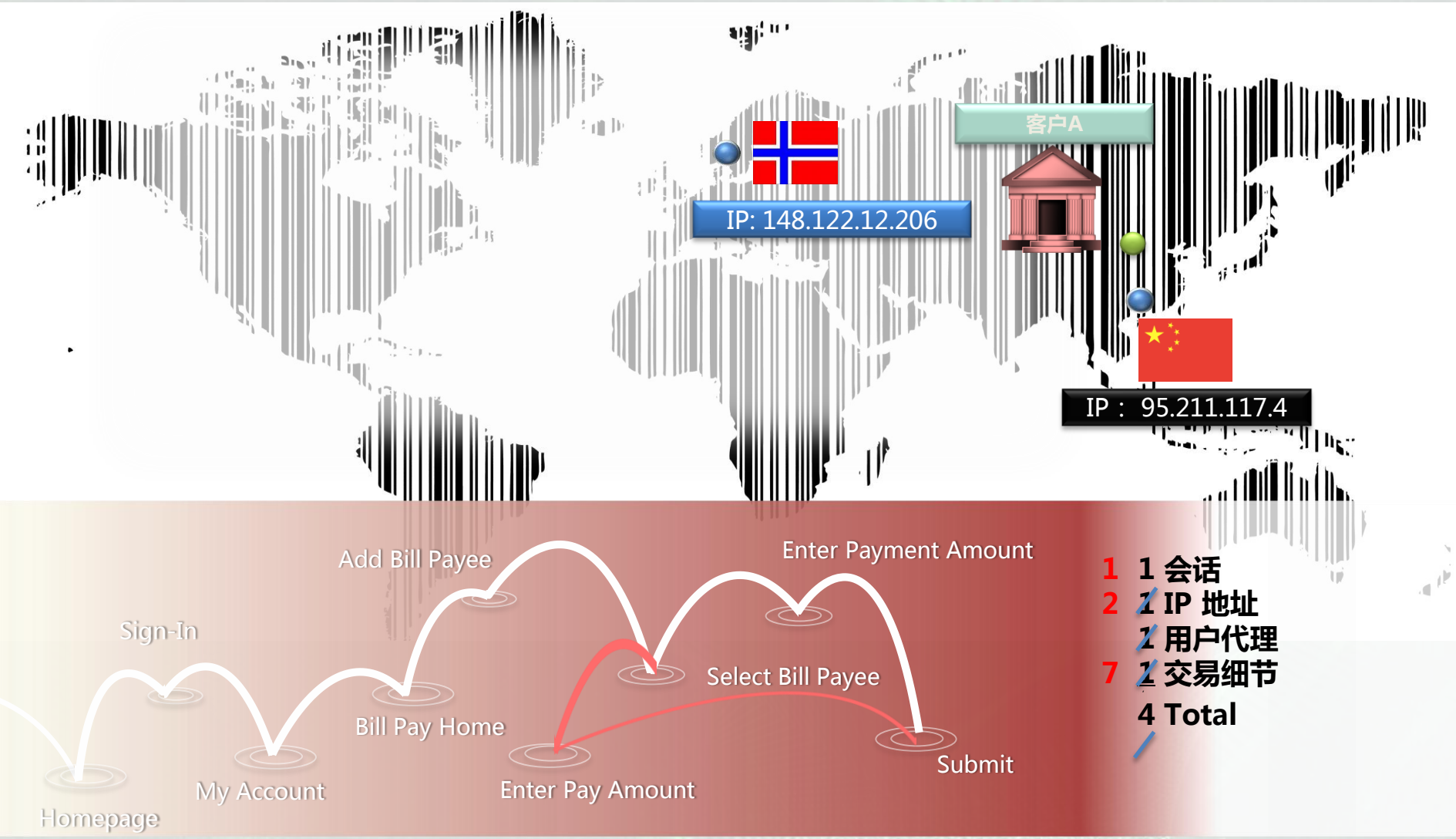
实时威胁组

- DDoS
- 自动化行为
- 被感染设备，如 BOTs

"The Mothership"


客户A
IP： 95.211.117.4
IP： 95.211.117.4
Homepage
Sign-In
My Account
Bill Pay Home
Add Bill Payee
Enter Pay Amount
Select Bill Payee
Enter Payment Amount
Submit

非正常网站访问！


客户A
IP: 148.122.12.206
IP： 95.211.117.4
Homepage
Sign-In
My Account
Bill Pay Home
Add Bill Payee
Enter Pay Amount
Select Bill Payee
Enter Payment Amount
Submit
1 1 会话
2 1 IP 地址
1 用户代理
7 1 交易细节
4 Total

谢　谢