移动视角下的网络行为初探

——网络资源评估

摘要：

1. 网络资源评估，侧重关注网络资源个体间的关联关系，进行群体性事件的挖掘。以期及早感知新增威胁，有效挖掘历史隐含的未知事件。
2. 网络资源评估，分为资源关联和能力评估两个阶段。
3. 资源关联阶段是在各独立的网络资源个体间构建起关联关系，寻找同类资源的过程。
   1. 同类资源关系的确认，将暴露隐含未知事件的更多线索。
   2. 同类资源的规模大小，辅助决策事件关注的优先级。
   3. 同类资源的时间维度来看，分为主动和被动处理；
   4. 主动处理，是针对新增网络资源，如果它与历史资源发生了关联，则优先进入观察视野，对同类资源进行含义解读、标识。
   5. 被动处理，是针对不再活跃的历史网络资源，集中进行事件挖掘的过程。
4. 能力评估阶段，建立在资源关联的基础之上，针对其背后的人、组织、团体的规模、行为特征、能力的评价；同时，也是寻求更为高效的查杀策略的过程。
5. 网络资源评估，相对于人的评估，更具有普适性、客观中立性，不致受舆论等因素限制。
6. 希望网络资源评估在移动安全领域，特别是样本分析、事件溯源跟踪中，能够得到业界同仁的充分重视，提升工程化效率。

# 概念

## 网络资源

网络资源，也叫[网络信息资源](http://baike.so.com/doc/649426-687383.html)，是指通过计算机网络可以利用的各种信息资源的总和。

具体到本文中，是指：

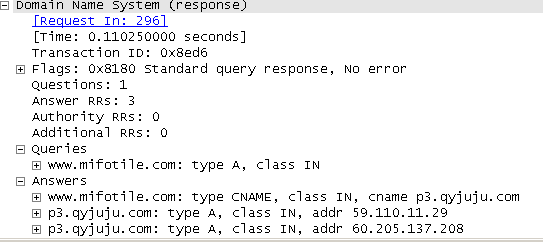
1. 以PC、移动端设备为代表的终端及运行其上的应用软件，在同外界通信、交互过程中，所涉及对象、产生的信息流等。
2. 包括但不限于
   1. 远端服务器信息： domain/host/ip/port

远端交互服务器，它的域名信息（domain），子域名分布情况（domain/hosts）

* 1. DNS信息：请求链、历史解析记录

关注DNS信息，不能仅关注其解析到的IP地址:从原始的请求，到最终的应答，这中间的递归信息记录，都有助于构建关联关系。

该点在CDN场景中有显著性体现。

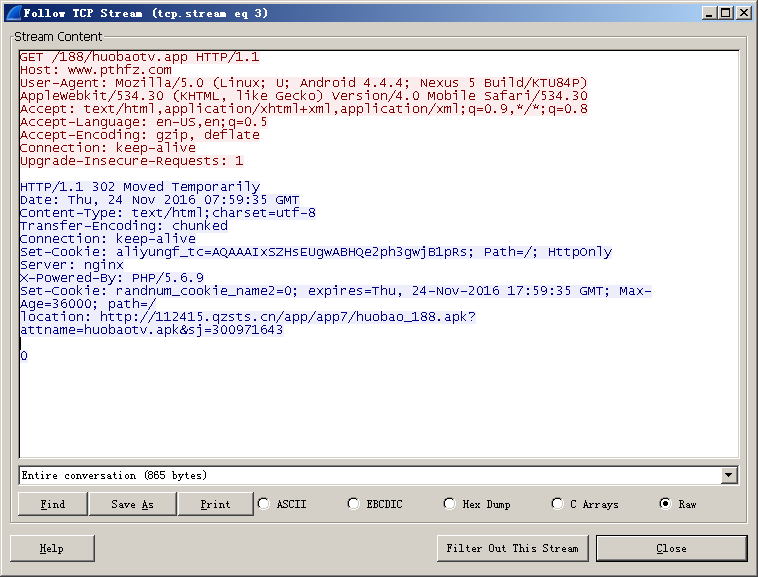


* 1. URL信息： 单url信息， 多url的请求链

url模式信息，代表资源的部署规律；

具备类似url模式的不同url间，其[背后存在着可能的关联](http://so.baiten.cn/detail/patentdetail/63/CN201610900534.9/20)。

以HTTP 30X为代表的不同url间的跳转关系，可帮助构建url间的关联。



* 1. 应用软件信息（非必须）：触发网络交互行为的主体

网络交互行为因样本主体而触发，但更多时候是无法获取主体信息的。

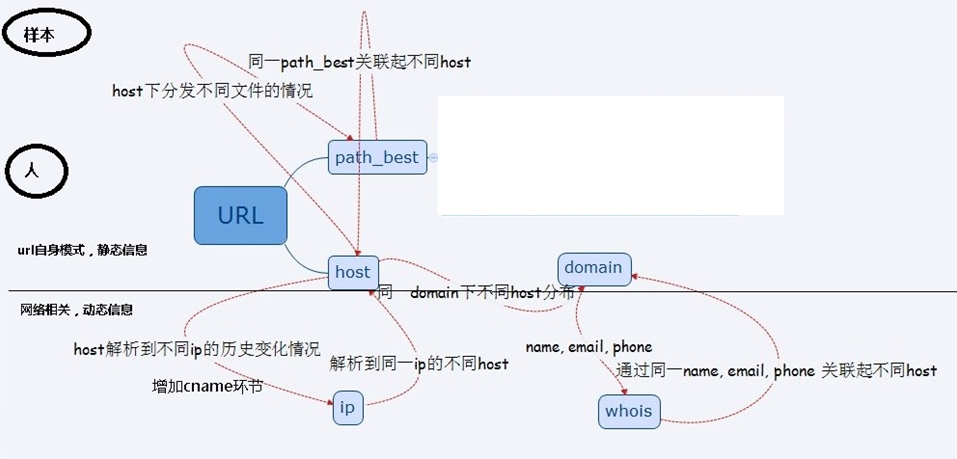
另外一个角度， 网络行为也不应该局限于特定的平台、个体，应该具有通用性。

因此，样本主体信息是非必须的。

* 1. whois注册信息(非必须)：

1. 特别的， 本文探讨的范围主要集中在移动端恶意样本的分析、事件的溯源跟踪领域。

上述网络资源，因背后拥有者的部署、驱动而产生、存在，可视为拥有者的属性。网络资源一定程度上反应了拥有者的特性。



## 网络资源评估

网络资源评估概念的阐述，依赖于以下基本概念：

1. 种子资源：或者种子信息， 是已知的某一个或者某一批特定的网络资源。
2. 同类资源：或者关联资源， 由于种种规则而与种子资源建立起关联的其他资源。其可能的分类：
3. 为同一个体、组织、团体所拥有：主要依据dns, whois
4. 为不同个体、组织、团体所拥有，但具备同套、或者类似的源码、部署等机制，视为同源性： 主要依据url模式
5. 已知资源的上游来源： 主要依据HTTP请求链，重定向信息
6. 已知资源的下游去向： 主要依据HTTP请求链，重定向信息

基于以上信息，我们对网络资源评估进行定义。

网络资源评估，包含两部分：

1. 资源关联

针对已知的种子资源，获取其同类资源，本质是资源关联的过程。是以少看多，看全局，看整体的过程。

1. 能力评估

针对获取的关联资源，通过对资源规模、模式特点的分析，增强样本识别能力，提升查杀效率；去评估其背后的个体、组织、团体的规模、行为特征、能力。

关于资源评估与人因素的评估的关系：

1. 资源附属于人
2. 人具备主观能动性，资源相对客观。
3. 资源评估是对人评估的一个部分，是其中一个视角。
4. 对人的评估变得复杂，且因人而异，不具备普适性。
5. 而资源评估具备中立、客观性，具备普适性，可以自动化、工程化。
6. 特别的，对人、团体的评估，有时会受舆论等因素的限制，不方便进行。此时，可转而评估其背后的网络资源，做出客观、中立的评估。

# 为什么要做网络资源评估？

网络资源评估，侧重关注网络资源个体间的关联关系，进行群体性事件的挖掘。以期及早感知新增威胁，有效挖掘历史隐含的未知事件。

网络资源评估，其意义在于：

1. 服务于溯源跟踪领域的线索收集，本质是一个信息收集、线索收集的过程。
2. 服务于背后人员、组织、团队的关联性分析

是否有共同的服务器，或者采用了同套的源码进行传播、部署。

1. 服务于背后的人员、组织、团体行为特征、能力的评估

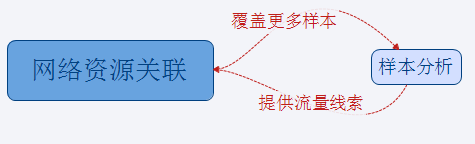
同类网络资源的规模，一定程度上可以反应背后人员、组织的资源占有、协调、传播能力。

1. 服务于背后人员、组织、团体的性质评价

当出于舆论等因素限制，无法针对人的因素给予评价时，可考虑转而客观、中立地评估其具有的同类网络资源。

此外，单从样本分析的角度，网络资源评估是样本分析的有效切入点：

1. 多数样本都是要与外界进行交互的。
2. 而与外界交互的过程，也是最易暴露的。
3. 网络流量是与外界交互的一个重要渠道。
4. 因此，对网络流量的侧重、有效分析，是提升样本分析效率的关键所在。
5. 反观目前移动端的样本分析领域， 注意力多集中在单样本、单域名内，少有进行背后整体的评估。
6. 希望以此“网络资源评估”概念，推进行业内分析步伐。



网络资源角度的查杀，可作为样本查杀体系的有效补充：

1. 样本利用加密、混淆等手段，以对抗现有的查杀体系。
2. 而网络资源角度的查杀，可跳脱样本代码的对抗维度，针对网络资源的规律，进行有效查杀。

# 如何做网络资源评估？

在前面提到，网络资源评估包含：

1. 资源关联

意在强调以少看多，侧重关注网络资源个体间的关联关系，进行群体性事件的挖掘。

1. 能力评估

是同类资源规模、能力的评估，也是背后人员、组织能力、行为特征的评估。

同时， 也提到对人评估的复杂性、不具可复制性以及可能的舆论因素限制。

因此，这里希望网络资源评估应该尽量做到以下几点：

1. 自动化、工程化、

网络资源评估成为日常分析工作中必不可少的一部分，决定了其需要具备自动化、工程化的特性，在信息收集、关联方面解放人力投入。

1. 可复制化

网络流量自身具有跨平台的特点，一个好的网络资源评估方法也应尽量做到具备可移植性、通用性，不应有太多的外界依赖。

比如，样本主体信息的获取在有些场景中是无法获取的，因此，样本主体在网络资源评估中是可有可无的，不应该成为一个限制。

再如，对于DNS历史解析记录， 也应该是优先维护所在自给自足小环境（如沙箱环境）的数据，不应该过多考虑外界DNS， 以造成噪音干扰。

相对的，对于注册信息whois更应该持观望态度，至少其不应该成为网络资源评估中不可缺少的一环。 特别是，受限于注册信息的隐私保护机制、租赁扎堆现象的存在，其产生的噪音往往多过其带来的有益价值。

1. 以少看多，尽量多的进行资源关联

对于给定的一批种子，在避免误报噪音的情况下，应该尽可能多的给出可能的关联资源、线索信息，为分析人员提供参考。

那么，具体如何做网络资源评估呢？

## 资源关联

针对已知的种子资源，获取其同类资源的过程，本质是资源关联的过程。

### 种子资源

种子资源可以是主动或者被动的一个、一批资源：

1. 被动—— 是针对既定已知事件的快速跟进过程
2. 主动—— 是针对日常分析、可疑性内容的主动挖掘、发现的过程

这些资源线索输入可以是：

1. url
   1. host部分： domain/host/ip/port
   2. 非host部分关键词: url模式
2. dns

dns解析链的各环节数据， 如A记录、CNAME记录等

1. http请求链

从发起请求，到最终应答的各环节，其都是可能存在关联、依赖的。

如HTTP 30X重定向信息

1. 样本hash

非必须维护

如果是样本，需首先获取其触发的流量信息，或者样本的下载来源。

根据这些网络流量信息，进行再扩展。

1. whois

非必须维护

辅助参考

### 寻找关联

判定关联的主要依据，包括但不限于：

1. 相同的url模式：

url模式，如同文件路径，代表着服务器上资源的部署规律。

类似的url模式，代表着同类资源。

两个不同的主机， 因为具备类似的url模式，[对同类资源进行传播而建立起关联。](http://so.baiten.cn/detail/patentdetail/63/CN201610900534.9/20)

1. DNS请求链中存在相同的环节
   1. 如解析到同一个IP
   2. 具备相同的CNAME记录
2. HTTP请求链中存在相同的环节
   1. 如HTTP 30X重定向中，存在相同的原始请求。
3. whois (辅助)

以上各资源虽然迥异，但可以整合到同一个数据结构中：

1. 同一个数据结构的维护，可以在不同资源类型间建立起关联，通过直接或者间接的关联关系。
2. 同一个数据结构的维护，更易于自动化实现。

## 3.2 后分析 （Post-Analysis）

上述资源关联环节，可以找到一批同类的网络资源。

针对这些同类网络资源的进一步分析，可以进一步揭示其现象特点，以及背后可能人员、组织的能力。

1. 这些客观、同类网络资源的规模如何。

相对于其他其他同类网络资源，其表现是否突出，是否值得特别跟进。

1. 这些同类资源间是否存在其他角度的共性，以便于查杀策略的提取、落地。
2. 从同类网络资源转而分析其背后的部署者、人员信息
   1. 背后人员的能力评估，如开发能力、资源协调能力
   2. 背后人员的行为特征，如编码习惯、命名习惯
   3. 背后人员、组织的规模



# 案例