比特币 P2P 网络中客户端的去匿名

Title： Deanonymisation of Clients in Bitcoin P2P Network

Authors： Alex Biryukov, Dmitry Khovratovich, Ivan Pustogarov

Institution： University of Luxembourg

Journal： CCS' 14

Topic： 比特币；隐私保护；网络通信

Motivation： 如何连结比特币地址和实际身份？

当用户在 NAT 或 ISP 的防火墙后的时候，去匿名比特币用户，即区分在同一个 NAT 后的不同用户的网络连接和交易，链接用户假名与发起交易的 IP 地址。

Approach： 在 Kaminsky 提出“第一个告诉你交易的节点可能就是交易的始发节点”的算法上进行优化，核心思路是始发节点转发交易后，始发节点的邻居节点会是第 2 批转发交易的节点，因此可以利用邻居节点转发信息的时间排序来定位始发节点。作者将探针搜集的交易信息按照到达时间分类排序，然后针对每一类交易抽取排名前 8 的节点和 8 个邻居节点进行比对，如果重合率超过阈值，即认为节点是交易的始发节点。这种方法扩展了判断依据，能够减少网络延迟等干扰条件的影响，有效提高推测的准确率。此外，这种攻击方法可以发现隐藏在 NAT 服务后面的客户端节点。

Phase： 1) 如何禁止比特币服务端接受通过洋葱路由（Tor）和其他匿名服务的连接？

Goal：当客户端链接到其他节点时，使用真实 IP 地址。

Idea：利用比特币的 DoS 保护机制。

当节点接收到畸形的消息时，它增加消息来源的 IP 地址的惩罚评分（如果客户端使用洋葱路由，那么消息来源是洋葱路由出口节点）。当评分超过 100 时，发送者的 IP 将被禁止 24 小时。也就是说，如果客户端通过洋葱路由中继代理他的连接并且发送畸形的消息，中继的 IP 的地址将会被禁止。

Phase： 2) 如何学习比特币客户端连接的入口节点？

当客户端C与一个入口节点建立时，它广播它的地址C_a。如果攻击者已经连接到该入口节点，那么有概率（取决于攻击者的连接数）C_a会被转发给他。

Idea：(i) 连接到W个比特币服务端，其中W接近于服务端的总数；

(ii) 对每个广播的C_a，记录转发C_a给攻击者的机器的服务端组E'，指定为入口节点子集E'_{C_a}。

Problem：(i) 入口节点可能将客户端的地址发送给某些非攻击者的节点；

(ii) 客户端不会同时连接到所有入口节点，连接之间有时间间隙。

在上述两种情况中，广播的地址通过非入口节点到达攻击者的机器，E'_{C_a}中产生假（噪声）入口。

噪声缩减技术

Hypothesis：(i) 客户端在 NAT 后，他的 IP 已经在比特币网络中使用；

(ii) 客户端的公共 IP 包含在主 ISP 的已知 IP 地址的列表中。

Fact：　如果地址已经从A发送给B，那么它不会在该连接上再次转发。

Idea：　当攻击者广播C_a时，每个比特币服务端选择两个责任节点转发地址。攻击者与每个服务端建立连接，希望他的节点替换部分转发C_a的责任节点。当客户端C连接到其中一个入口节点$e1$时，他广播他的地址。如果攻击者的一个节点替换了一个责任节点，那么攻击者将会学习到客户端C可能连接到节点$e1$。如果责任节点没有变更，那么地址C_a将不会在网络中被传播。

Phase：　3)　如何识别交易的发起者？

Step：　(i)　获取服务端列表S。该列表是定期刷新的：
攻击者手机通过$GETADDR$消息查询所有已知节点，收集全部节点的列表。对于每个在响应$ADDR$消息中的地址P，与它建立 TCP 连接并发送$VERSION$消息，如果它在线，那么P是服务端。

Step：　(ii)　生成去匿名化的比特币客户端列表C：
攻击者选择一组想要揭露其身份的节点C。

Step：　(iii)　当客户端C连接到网络时，把他们映射到各自的入口节点E_P：
运行第二阶段的程序。攻击者实际获取E'_P。

Step：　(iv)　监听服务端S，把交易映射到入口节点：
攻击者监听所有建立的连接的$INVENTORY$消息和接收的交易的哈希值，对于每条交易T，攻击者收集前q个转发$INVENTORY$消息的比特币服务端的地址R_T。对比E'_P和R_T，匹配入口建议对(P,T)。

- 攻击者在R_T查找所有组E'_P的出现情况，生成所有可能的三元组。如果有匹配，获取对(P,T)。
- 如果没有匹配，攻击者考虑二元组，然后一元组。在后续的交易中过滤几对建议的(P,T)。

Contribution：　去匿名大量比特币用户并关联它们的假名和公共 IP 地址的通用方法：

1)　把客户端作为攻击目标，区分有相同公共 IP 的节点；

2)　客户端使用如 Tor 等匿名服务；

Performance：　实验显示识别准确率为 11%，如果采用一些辅助的攻击，准确率能提高到 60%。

Dataset：　比特币的测试网络

Baseline：　/

Metric：　1）　入口节点是否第一批转发交易？

2）　执行交易的所有步骤？