Summary of experiment platform requirements

Goal

We want to capture data that resembles behaviour of criminal proxies. These proxies are relaying services and content reserved to customers/employees to a wider audience. Examples we discussed are:

- DNS resolvers forwarding IPS not accessible to the public.
- Proxies relaying video streams from plattforms such as BT Sports to the public.

The ultimate goal here is to detect these proxies on a wider scale from the perspective of an ISP. Apart from the obvious incentive to stop restricted content being leaked to the public, another motivation here is the assumption that these proxies belong to a larger criminal infrastructure and are often reused for other malicious operations.

Data capture experiment

Our goal is to capture data resembling behaviour of such proxies by setting up a proxy that forwards content from BT sports to selected hosts.

Data capture constraints

ISP perspective

The main constraint for this project is the observation of traffic from an ISP perspective, i.e. we can only record every n-th packet. To my knowledge, there are also constraints on the creation of flow summaries and not all existing flows are recorded.

An easy way to ensure the perspective constraint is met is by recording the traffic with an actually deployed probe.

contact host requirements

Relay proxies are supposedly contacted by many different hosts that all request similar content.

The number of hosts contacting the proxy should be quite high to resemble realistic behaviour, as this is an important feature

Content requirements

Depending on the assumption we make, the number of hosts the proxy relays requests to (i.e. the servers providing protected content) should probably be a small to moderate number. We did not discuss yet if such proxies are likely used for more than one purpose.

Forwarding requirements

The type of content each of the targeted servers provides may vary, i.e. for BT Sports it might be both live video (via UDP?) as well as regular video (TCP/HTTP). We have to account for the fact that the used traffic protocols might as well

Network settings

I think the realism of the network setting in terms of bandwidth, packet loss etc. is negligible since to my knowledge we do not capture interarrival distributions etc.

To our knowledge, the IP address of the proxies is quite sticky, might only change once a day. Not really worth considering this constraint.

Open questions

- How narrow are proxies in their scope? Are they restricted to a specific set forwarded services, i.e. one proxy exclusively for BT Sports, or are they set-up t provide many different services? Could one proxy even have multiple functionalities, such as one proxy acting simultaneously as a DNS resolver and a video forwarder?
- How are these proxies advertised/contacted? Do users need specific implementations/software, or specific configurations? Or are they contacted via existing websites? How does the "user" actually use them? Does this vary a lot?