I2P's Usage Characterization

Juan Pablo Timpanaro, Isabelle Chrisment, Olivier Festor

I2P's Usage Characterization

Juan Pablo Timpanaro, Isabelle Chrisment*, Olivier Festor

INRIA Nancy-Grand Est, France
*LORIA - ESIAL, Henri Poincaré University, Nancy 1, France

Abstract. We present the first monitoring study aiming to characterize the usage of the I2P network, a low-latency anonymous network based on garlic routing. We design a distributed monitoring architecture for the I2P network and show through three one-week measurement experiments the ability of the system to identify a significant number of all running applications, among web servers and file-sharing clients.

1 Introduction

The I2P network[1], mainly designed to allow a fully anonymous conversation between two parties inside the network, contains a full range of available applications. The goal of this work is to characterize the usage of the I2P network and thereby to answer the following question: what is the I2P network used for?. We consider that this analysis is important for the network improvements. By determining that most of I2P users are, for example, file-sharers, we could enhance the network in that direction.

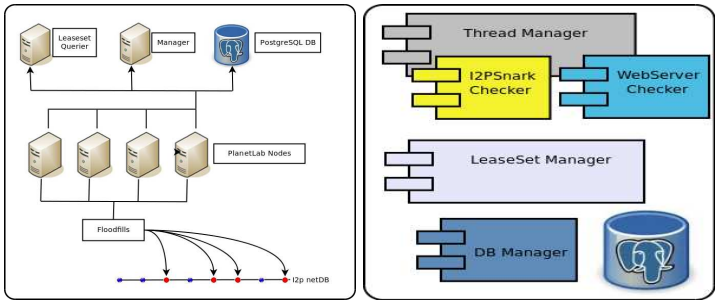
To do that, we propose a fully operational architecture to monitor the I2P network at application-level. Then, we show we can identify which applications are most used and when they are used, as opposed to statistics web sites[2] giving only the number of applications, but not the type.

2 I2P monitoring architecture

In the I2P network, a distributed hash table based on the Kademlia [**?**] protocol is used to store and share network metadata. However, contrary to the Kademlia protocol, only the fast I2P users form part of the DHT. They are called the *floodfill* peers. There are two types of network metadata: *leasesets* and *router infos*. A leaseset provides information about a specific destination, like a web server, a BitTorrent client, an e-mail server,.... A router info provides information about a specific router and how to contact it, including the router identity (keys and a certificate), the address to contact it (ip and port) and a signature.

Our monitoring architecture is divided in two main parts:

[1] `http://www.i2p2.de/`
[2] Such as `http://stats.i2p.to/`



(a) Leasesets retrieval (b) Leaseset querier

Fig. 1. Complete monitoring architecture

- A first part is responsible for retrieving and collecting *leasesets* as described in Figure 1(a). Each floodfill is running in one PlanetLab node, and logs every received request to a PostgreSQL server, located in our security lab[3].
- A second part, the *leaseset querier*, runs in parallel and is shown in Figure 1(b). The *Leaseset Manager* periodically retrieves new leasesets from the database, while the *Thread Manager* creates different threads to test a given destination for a particular application (a Web Server or an I2PSnark client). Once a leaseset is tested, the result is kept in the PostgreSQL server.

We focus mainly our study on two kinds of applications running on top of an I2P router and published in the DHT: a web server and an I2PSnark client, which is a modified BitTorrent client.

To tag a given leaseset as a web server, we open an I2P connexion through it and send a `GET` message. If the response contains well-known http keywords, then that leaseset corresponds to a web server.

To test an I2PSnark client, we use the following approach. Once the I2P connection is established, we send a first message, a well-formed BitTorrent message, requesting a random torrent. If that given leaseset is actually running an I2PSnark client and not sharing the torrent, it will immediately close the socket. Secondly, we re-open the connection and send a malformed BitTorrent message. If the response timeouts, then we conclude that the given leaseset is running an I2PSnark client. If we do not receive any answer for the first message, we can not assume anything.

3 Experiments

3.1 Setup

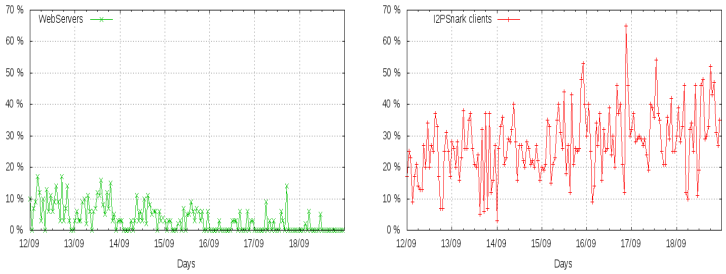
During these first series of experiments we used the Planet Lab test-bed. Planet Lab nodes have restrictions regarding the available bandwidth, and therefore we placed 15 floodfills with minimum bandwidth settings, still allowing us to

[3] `http://www.inria.fr/actualite/mediacenter/inauguration-lhs-nancy`.

perform as floodfills nodes. Our leaseset querier ran in parallel, analysing new leasesets and storing the results. We conducted 3 different one-week experiments, starting on September 12th, September 27th and October 4th. We logged four possible outcomes for a given leaseset: *WebServer*, *I2PSnark*, *Unknown* (The given leaseset is not running a web server nor an I2PSnark client) and *Destination not found* (The given leaseset is unreachable).

3.2 First results

Figure 2 gives the percentage of leasesets that we were able to identify and we tagged for the first experiment. In average, we identified 32.06% of all the leasesets queried, with most of them being I2PSnark clients, as shown in figure 2(b).



(a) Anonymous web servers (b) I2PSnark clients

Fig. 2. Identifiable leasesets (September 12th - 19th)

Table 3 summarizes the number of I2P destinations we analyzed for every experiment. For our last experiment, the *Destination not found* value is higher than we saw in the previous experiments. The statistics web page of I2P[4] reports that between the 10/02 and the 10/09, the number of routers in the network increased considerably (from 7000 routes to 10000 routers.) This increase of new participants in the network could explain why we have a higher number of unreachable destinations, since many I2P users might have been trying the network and testing different applications.

	Anonymous web servers	I2PSnark clients	Unknown	Destination not found	Total
First experiment	193	1312	1423	1793	4721
Second experiment	172	1106	1503	1757	4538
Third experiment	175	1057	1186	2349	4767

Fig. 3. Results for every one-week experiment

We have, on average, 30.16% of *unknown* applications, which means that we successfully opened a connexion through these leasesets, but we failed at tagging them.

[4] `http://stats.i2p.to/`

4 Related work

McCoy et al.[**?**] analysed the application layer of outgoing traffic in Tor, to determinate the protocol distribution in the network. This study resembles what we want to achieve for I2P, however, the data collection methodology applied by McCoy et al. can not be used in I2P. More recently, Loesing et al. [**?**] presented a measurement of sensitive data on the Tor network, such as countries of connections and outbound traffic by port. Additionally, Loesing [**?**] measured the *trends* of the Tor network from directory information. Nevertheless, this network has a central component, a *directory server* and hence the monitoring approaches can not be applied to the I2P network.

5 Conclusion

This study shows a series of experiments on the I2P network. After the first one-week experiment, and after analysing 4721 leasesets, we were able to identify 193 web servers and 1312 I2PSnark clients, and we determined that 37% of the published leasesets were off-line after their publication on the netDB.

Monitoring a network is crucial to understand what it is used for. We consider it a mistake to apply previous well-known results in p2p networks, like most users performing file-sharing during weekends, in anonymous networks before classifying its traffic accordingly. Anonymous networks allow a user to keep its identity safe while web-surfing or file-sharing, for example. Moreover, an anonymous network like I2P, which provides a set of built-in services, can be used for different objectives and in different manners when comparing it to non-anonymous networks, or even to anonymous networks, like *Tor*.

This work is the first step on monitoring and classifying the traffic on the I2P network, in order to understand the network. A full version of this work can be found in [**?**]. Other experiments and measurements are in progress.

Acknowledgment: We thank the anonymous leading developer of I2P for his useful comments and reviews of this paper.