Monitoring Anonymous P2P File-Sharing Systems

Juan Pablo Timpanaro, Isabelle Chrisment, Olivier Festor

Monitoring Anonymous P2P File-Sharing Systems

Juan Pablo Timpanaro
INRIA Nancy-Grand Est, France
Villers-lès-Nancy
F-54600, France
Email: juanpablo.timpanaro@inria.fr

Isabelle Chrisment
TELECOM Nancy, Université de Lorraine
LORIA UMR 7503
Vandœuvre-lès-Nancy
F-54602, France
Email: isabelle.chrisment@loria.fr

Olivier Festor
INRIA Nancy-Grand Est, France
Villers-lès-Nancy
F-54600, France
Email: olivier.festor@inria.fr

***Index Terms*—Distributed monitoring, Large scale monitoring, I2P, anonymous networks**

I. Motivation

Anonymous communications have been exponentially growing [1], [2], where more and more users are shifting to a *privacy-preserving* Internet and anonymising their peer-to-peer communications [3]. Anonymous systems allow users to access different services while preserving their anonymity. We aim to characterise these anonymous systems, with a special focus in the I2P network [4].

Current statistics service for the I2P network [1] do not provide values about the type of applications deployed in the network nor the geographical localisation of users. Our objective is to determine the number of users in the network, the number of anonymous applications, and the type of those applications. We also explore the possibility of inferring which group of users is responsible for the activity of an anonymous application. Thus, we improve the current I2P statistics and get better insights of the network.

II. I2P monitoring approach

The information about an I2P node[1] is grouped in a structure called *routerinfo*, which includes its contact information (IP address and port), encryption keys, among others. The information regarding an I2P application is gathered in a structure called *leaseset*. Therefore, an I2P user running two anonymous applications has one routerinfo and two leasesets.

The I2P network guarantees the *unlinkability* between a routerinfo (defining a user) and a leaseset (defining an applications). We propose a correlation analysis between the behaviour of users and applications. This analysis can allow us to assert in which measure a particular set of users contributes to the overall activity of an anonymous applications.

The I2P network uses a Kademlia-based distributed hash table, called *netDB*, to store network metadata, that is, every routerinfo and leaseset. This database is formed by normal I2P nodes with high bandwidth rates, called *floodfill nodes*. These I2P nodes, or floodfill nodes, participate in the network as every other node, and additionally store network metadata. In the current I2P network, these floodfill nodes account for nearly 2% of the total network (600 floodfill nodes out of 35000 I2P nodes).

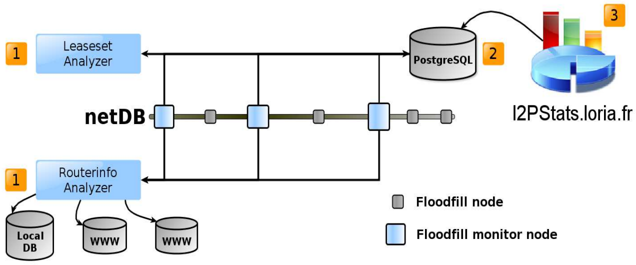

Fig. 1. Monitoring architecture

We exploit the netDB and place a set of *monitoring floodfill nodes*. These nodes allow us to collect a vast amount of network metadata, which is analysed to characterise I2P users (by analysing routerinfos) and I2P applications (by analysing leasesets).

Figure 1 depicts our monitoring architecture. Monitoring floodfill nodes are deployed in the I2P's netDB, analysing every routerinfo and leasesets they retrieved. The result of these analyses are stored, for a later correlation analysis. Finally, the aggregation of these results are displayed in our statistics web site[2].

Our monitoring architecture is completely flexible. Due to the autonomous nature of floodfill nodes, we can increase the number of monitoring floodfill nodes on demand, improving the amount of network metadata we receive, thus enhancing our monitoring results.

III. Experiments

This section presents the number of users and the number of anonymous file-sharing applications during our measurements. We additionally present a correlation analysis to determine in which extent the users from Moscow contribute to the I2P's anonymous file-sharing activity.

A. Experimental setup

We deployed our monitoring architecture for a period of one week, from 2013-03-15 CEST to 2013-03-21 CEST. We use seventy monitoring floodfill nodes, which in the current

[1] An I2P user runs an I2P node. We employ *I2P user* and *I2P node* indistinctively throughout this document.

[2] Accessible at www.i2pstats.loria.fr.

978-1-4799-0521-8/13/$31.00 ©2013 IEEE

netDB, account for nearly the 12% of the total number of floodfill nodes.

B. I2P's file-sharing applications

Figure 2 displays the individual number of active file-sharing applications during the duration of our analysis, for I2PSnark, IMule and I2Phex clients. We can observe that IMule and I2Phex clients hardly presented any activity, and I2PSnark was the most deployed file-sharing client. We can additionally observe that the use of I2PSnark clients increased during the weekend (from `16/03` to `17/03`).

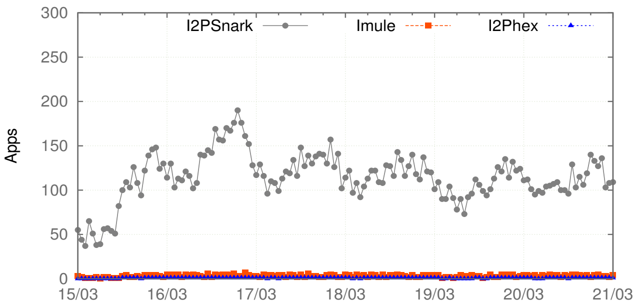

Fig. 2. Detection of I2P's file-sharing applications

C. Countries & Cities analysis

Figure 3 presents the most active countries in the I2P network: Russia, The United States, Germany and France. Russia was the most active country in the network, with an average of 3500 users.

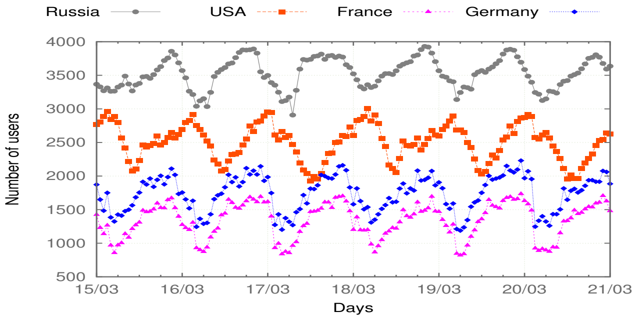

Fig. 3. Top four active countries

Figure 4 displays the top three cities. Moscow was most active city, with an average number of 710 users, while Saint Petersburg had an average number of 350 users. In both cases, country-based and city-based, Russia contributed the most.

During the entire one-week experiment, we detected 159 countries and 13547 cities. That indicates that the I2P network is widely deployed and distributed all over the world.

D. Correlation analysis

Our goal is to determine in which measure the users detected contributed to the anonymous file-sharing activity.

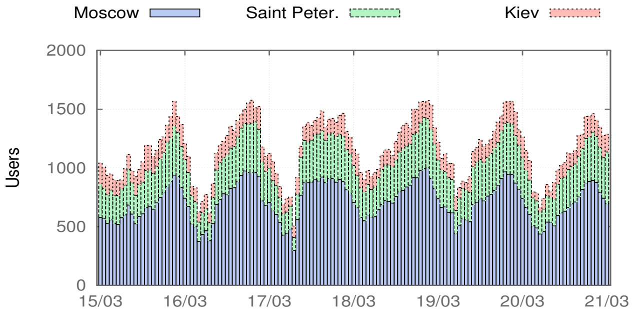

Fig. 4. Top active cities

We are dealing with data of type *ratio* [5], where Pearson's correlation coefficient is adequate. For the correlation analysis we extended our period of analysis on one week, from `2013-03-15 CEST` to `2013-03-30 CEST`. We consider that three weekends is a good time window to detect a long-lived correlation between a particular city and a particular application. We took into account the top detected city, Moscow, and the top detected file-sharing application, I2PSnark.

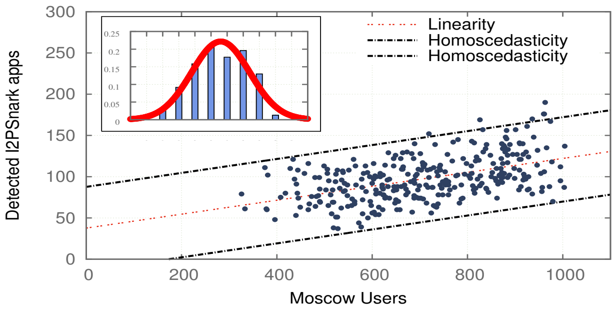

Fig. 5. Correlation analysis for Moscow

Figure 5 plots the data with the number of users from Moscow in the x-axis and the number of detected I2PSnark applications in the y-axis. The plotted data complies with Pearson's data requirements (data normality, linearity, and homoscedasticity). The correlation value is $r = 0.4901$ and indicates a strong correlation, where the coefficient of determination[3] is $r^2 = 0.2401$ and indicates that users from Moscow contributed to the 24% of the total I2P's file-sharing activity for the fifteen-day period.

References

[1] I2P. The home for NetDB statistics. http://stats.i2p.in/.

[2] Tor. Tor Metrics Portal. https://metrics.torproject.org/.

[3] Stevens Le Blond, Pere Manils, Abdelberi Chaabane, Mohamed Ali Kaafar, Claude Castelluccia, Arnaud Legout, and Walid Dabbous. One bad apple spoils the bunch: exploiting P2P applications to trace and profile Tor users. In *Proceedings of the 4th USENIX conference on Large-scale exploits and emergent threats*, LEET '11, Boston, MA, March 2011. USENIX Association.

[4] I2P. The I2P network. http://www.i2p2.de/.

[5] Stanley Stevens. On the Theory of Scales of Measurement. *Science*, 103(2684):677–680, 1946.

[3] The coefficient of determination explains the variance of the data, *i.e.* how much of the variance of the y-axis can be explained from the variance of the x-axis.