SSN project proposal

Finding the true identity of hidden webservers within I2P.

Iwan Hoogendoorn - iwan.hoogendoorn@os3.nl
Tarik El Yassem - tarik.elyassem@os3.nl
Joris Soeurt - joris.soeurt@os3.nl

November 17, 2011

1 Introduction

Several privacy enhancing overlay networks exist. Beside the largest and most popular network Tor, the I2P network is also relatively large. The major design difference with Tor is that I2P mainly focusses on hosting services within the network instead of creating an anonymous gateway to regular internet services. Adrian Crenshaw has written a paper[1] and proof of concept of how to link these hidden anonymous services to real world (non anonymous) IP addresses. Although the paper is of good quality we think there is room for improvement of various aspects. Timpanaro, Chrisment and Festor have researched[2] monitoring of the I2P network and have found that webservers hosted within I2P remain at a specific adress longer than servers that offer bittorrent services. Therefore it is more practical to focus on HTTP services within I2P for our research.

2 Research question

Main research question:

> "Is it possible to increase the percentage of I2P hosted websites that can be linked to real world IP addresses by improving techniques and using yet unused techniques?"

Sub questions:

- Is it possible to improve the completeness of the list of I2P hosted services?
- Is it possible to improve the completeness of the list of IPs of I2P using hosts?
- Is it possible to increase the number of matches between IP and I2P hosts?
- Is it possible to enhance the matches using other techniques than Adrian used so far?

3 Scope

This project is limited by the following scope:

- This project is purely based on the I2P network
- We will only try to identify services based on HTTP protocol
- We are not allowed to run an exit node for this research

4 Approach

We will try to answer the research question using the following approach:

- Read about the I2P network and protocol
- Reproduce the relevant parts of Adrian Crenshaw's paper
- Define which methods we will investigate
  - Current idea's include using meta information from files downloaded from websites, recursive crawling, using other sources for the I2P host list like address books, forums and search engines and using information gained from hosted webapplications.
- Use these methods to improve the accuracy
- Measure and compare the results
- Reach a conclusion, make suggestions for improvement and further research

5 Planning

We have defined the following actions:

Action	Person	Time
Initial research	ALL	18 hours
Create proposal	Joris and Tarik	8 hours
Create planning	Tarik	2 hours
Read about I2P network and protocol	ALL	8 hours
Setup base environment	Joris	4 hours
Reproduce Crenshaw's environment	Joris	4 hours
Reproduce Crenshaw's results	Tarik	8 hours
Define improvement methods	ALL	8 hours
Implement improvement method	ALL	*
Review result	ALL	*
Review whole of improvements	ALL	*
Draw general conclusions	ALL	*
Suggest further work	ALL	*
Write report	ALL	*
Create presentation	Iwan	8 hours
TOTAL	ALL	150 hours

6 References

References

[1] Adrian Crenshaw: "Darknets and hidden servers: Identifying the true IP/network identity of I2P service hosts"

[2] Juan Pablo Timpanaro, Isabelle Chrisment, Olivier Festor: "Monitoring the I2P network"