INFLOW: Inverse Network Flow Watermarking for Detecting Hidden Servers

Researcher

Dr. Alfonso Iacovazzi
alfonso_iacovazzi@sutd.edu.sg

sanat Sarda
sarda@sutd.edu.sg;

Prof. Yuval Elovici
elovici@inter.net.il

Publications

A. Iacovazzi, S. Sarda, D. Frassinelli, and Y. Elovici, "INFLOW: Inverse Network Flow Watermarking for Detecting Hidden Servers," IEEE-INFOCOM 2018, Honolulu, 2018.

Goals

Tor is a well-known and established communication system which allows its users to browse and communicate anonymously with fully guaranteed privacy, confidentiality, and content/service accessibility. When a content spreader needs to offer its content without being identified, they can use the hidden service system provided by the TOR network. However, this service has been increasingly abused, by distributing and hosting content, in most cases graphic, that are illegal or morally deplorable (e.g., child pornography). Law enforcement are continuously searching for means of identifying users and providers of such services. State-of-the-art techniques to breach the TOR anonymity are usually based on passive and active network traffic analysis, controlling TOR edge communication and exploiting TOR inherent flow transfer mechanism. Nonetheless, detecting hidden servers and linking illegal contents with the spreaders is still a challenging task that has not been completely accomplished.

Description

In this project, we describe INFLOW, a new technique to identify hidden servers based on inverse flow watermarking. INFLOW exploits the influence of congestion mechanisms on the traffic passing through the TOR network. INFLOW drops bursts of packets for short time intervals at the receiving side of a traffic flow coming from a hidden server and passing through the TOR network. Packet dropping affects the TOR flow control and causes time gaps in flows observed at the hidden server side. By controlling the communication edges and detecting the watermarked gaps, INFLOW is able to locate the hidden server. Our results, obtained by means of experiments performed on the real TOR network, show true positive rates in the range 90-98%.