Autonomous spambot detection

Ladislav Macoun, Tomáš Čejka
Affiliation (Faculty of information technology, CTU)
Thákurova 9, 160 00 Praha 6

macoulad@fit.cvut.cz

Keywords. network, monitoring, analysis, spam, bot, detection.

Abstract

It may have seem that these days, the emails are not as popular as the new modern instant messaging apps (IM) or the social networks. However, we still rely on an old SMTP protocol, which is the standard protocol for sending e-mails from one Message Transfer Agent (MTA) to another. The SMTP stands for "Simple Message Transfer Protocol", and it has been used since the early 90's, when the first RFC (RFC 821) was created. The creators have made a rock solid protocol which is still used for its reliability. Although its simplicity was a tradeoff for a security. The simplicity was abused by the spammers, at that time it was not a big deal as today. With the e-commerce boom had came many new businesses, scamming methods, and problems. One of them is the spam which we have to face today in enormous quantities.

The main goal of this project is not to detect spam such as its, but to detect entities and clusters that are sending it. All this with just a flow data analysis without interfering a network user privacy. This system contains two modules. First a static module that will autonomously decide whether the entity is a sender, legit mail server or potencial spammer. The second one is a clustering extension which will finds similarity between these potencial spammers using a SMTP flow header extension and create a clusters of them.

Acknowledgment

This was supported by the CTU grant No. SGS17/212/OHK3/3T/18 funded by internal grant of CTU of Prague.