Collecting and Analyzing Spam E-mails Through Simulating an Open Mail Relay

Degree programme: Master of Science in Engineering
Thesis advisor: Prof. Dr. Bruce Nikkel
Expert: Sylvain Hirsch

Spam e-mails are an annoyance to many people in the 21st century. To better understand and analyze, a service to collect large amounts of spam e-mails was designed and implemented. Compared to traditional spamtraps, this system actively coerces spammers into sending e-mails, therefore collecting a greater volume of messages.

Introduction

In the online world, spam emails are a major annoyance and security issue for regular users. Kaspersky estimates that more than 55% of global email traffic is unsolicited email. The most common way to study how spammers acquire email addresses and use them in their spam campaigns is to use so-called spamtrap addresses - email addresses specifically designed to receive spam. This project took a different approach: instead of spamtrap addresses, an SMTP server was implemented that simulates an open mail relay. It accepts any email for any address, while pretending to deliver it to the destination.

Implementation

A virtual machine in the BFH Cyber Lab with public IP addresses was provided to handle the incoming traffic. The design of the service is based on a microservice architecture using Docker containers. Each microservice has its own responsibility for the different topics. This allows parts of the system to be changed independently, while the rest can continue to operate unaffected. The main parts consist of the mail server container, which receives all incoming messages and stores them in the local file system. As a secondary step, the received emails are analysed and enriched with additional data in a separate container and finally stored in the database. Using the web interface and API, users of the system can browse the acquired dataset and search for specific domains or IP addresses. For deeper analysis, a JupyterLab instance is also provided to programmatically interact with the data.

Result

After the initial test emails were forwarded to the spammers' mailboxes, the number of spam emails increased dramatically. In total, over 180 million emails were received during the project. This shows that spammers are still actively looking for ways to abuse open mail relay servers. The dataset shows that more than 80 percent of the spam emails received were sent from countries in Europe, with the Netherlands being the top destination followed by France and Germany. This suggests that spammers have a bias towards choosing an Open Mail Relay closer to their own location.

Jonas Liechti

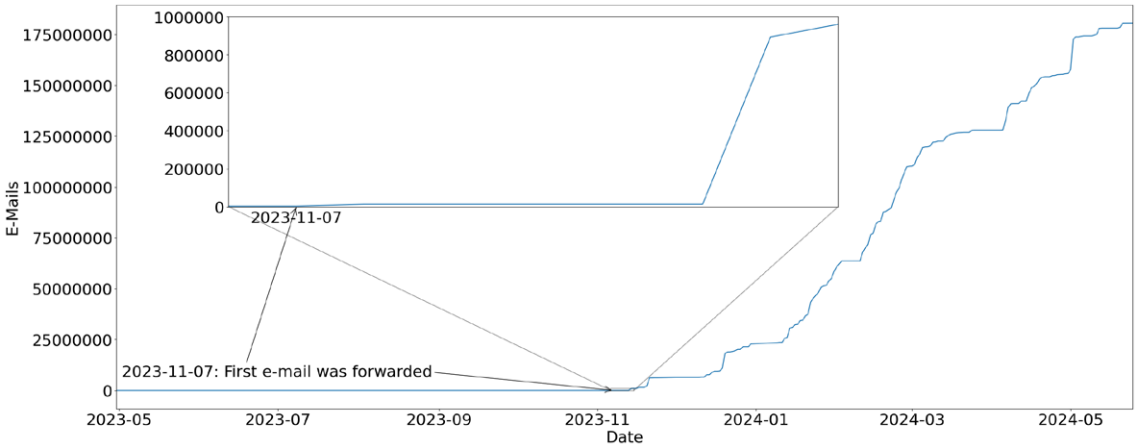

Figure 1: Number of received e-mails increased after actively forwarding the first test e-mails