Abstract

In recent years, cyber security incidents have occurred frequently. In most of these incidents, attackers have used different type of spam email as a knock-on to successfully invade government systems, well-known companies, and websites of politicians and social organizations in many countries. The detection of spam mail from big email data has been paid public attention. However, the camouflage technology of spam mail is becoming more and more complex, and the existing detection methods are unable to confront with the increasingly complex deception methods and the growing number of email. In this project, we proposed to design a novel efficient approach named Spam Spoiler for big e-mail data classification into four different classes: Normal, Fraudulent, Harassment, and Suspicious E-mails by using LSTM based GRU. The new method includes two important stages, sample expansion stage and testing stage under sufficient samples. The LSTM based GRU efficiently captures meaningful information from E-mails that can be used for forensic analysis as evidence. Experimental results revealed that Spam Spoiler performed better than existing ML algorithms and achieved a classification accuracy of 98% using the novel technique of LSTM with recurrent gradient units as there are different types of topics are discussed in E-mail content analysis. Spam Spoiler effectively outperforms existing methods while keeping the classification process robust and reliable. It utilizes Python, since it's a web application that requires data analysis and machine learning capabilities. Python can be used to build the backend of the application, as well as to perform data analysis and machine learning tasks.