ITPP0641
ICT

ARC'14
مؤتمر مؤسسة قطر السنوي للبحوث
QATAR FOUNDATION ANNUAL RESEARCH CONFERENCE
Towards World-class Research and Innovation

Email Authorship Attribution In Cyber Forensics

10.5339/qfarc.2014.ITPP0641

Rachid Hadjidj

Corresponding author :
rhadjidj@qu.edu.qa
Qatar University, Doha, Qatar

Abstract

E-mail is one of the most widely used forms of written communication over the Internet, and its use has increased tremendously for both personal and professional purposes. The increase in e-mail traffic comes also with an increase in the use of e-mails for illegitimate purposes to commit all sort of crimes. Phishing, spamming, e-mail bombing, threatening, cyber bullying, racial vilification, child pornography, viruses and malware propagation, and sexual harassments are common examples of e-mail abuses. Terrorist groups and criminal gangs are also using e-mail systems as a safe channel for their communication.

The alarming increase in the number of cybercrime incidents using e-mail is mostly due to the fact that e-mail can be easily anonymized. The problem of e-mail authorship attribution is to identify the most plausible author of an anonymous e-mail from a group of potential suspects. Most previous contributions employed a traditional classification approach, such as decision tree and Support Vector Machine (SVM), to identify the author and studied the effects of different writing style features on the classification accuracy. However, little attention has been given on ensuring the quality of the evidence. In this work, we introduce an innovative data mining method to capture the write-print of every suspect and model it as combinations of features that occur frequently in the suspect's e-mails. This notion is called frequent pattern, which has proven to be effective in many data mining applications, but has not been applied to the problem of authorship attribution. Unlike traditional approaches, the extracted write-print by our method is unique among the suspects and, therefore, provides convincing and credible evidence for presenting it in a court of law.

Experiments on real-life e-mails suggest that the proposed method can effectively identify the author and the results are supported by a strong evidence.