2016 International Conference on Computer and Communication Engineering (ICCCE)

Identification and Evaluation of Discriminative Lexical Features of Malware URL for Real-Time Classification

Morufu Olalere; Mohd Taufik Abdullah; Ramlan Mahmod; Azizol Abdullah

Abstract:
This study identifies and evaluates discriminative lexical features of malware URLs for building a real-time malware URL classification. The lexical features of malware URL are first identified from existing blacklisted malware URLs through manual examination. Feature identification is followed by studying the prevalence of these features on newly collected malware URLs through empirical analysis. Our empirical analysis revealed that attackers follow the same pattern in crafting malware URL. To evaluate the performance and effectiveness of these features, we applied a Support Vector Machine (SVM) classification algorithm on a dataset comprising of benign and malware URLs. By applying the WEKA data mining tool on our trained dataset, a 96.95 % accuracy was achieved with a low False Negative Rate (FNR) of 0.018 and a moderate False Positive Rate (FPR) of 0.046.

Published in: 2016 International Conference on Computer and Communication Engineering (ICCCE)

Date of Conference: 26-27 July 2016

Date Added to IEEE *Xplore*: 09 January 2017

ISBN Information:

INSPEC Accession Number: 16583701

DOI: 10.1109/ICCCE.2016.31

Publisher: IEEE

Conference Location: Kuala Lumpur, Malaysia

https://ieeexplore.ieee.org/document/7808289/keywords#keywords