

The phishing is a technique used by cyber-criminals to impersonate legitimate web-sites in order to obtain personal information. This paper presents a novel lightweight phishing detection approach completely based on the URL (uniform resource locator). The mentioned system produces a very satisfying recognition rate which is 95.80%. This system, is an SVM (support vector machine) tested on a 2000 records data-set consisting of 1000 legitimate and 1000 phishing URLs records. In the literature, several works tackled the phishing attack. However those systems are not optimal to smartphones and other embed devices because of their complex computing and their high battery usage. The proposed system uses only six URL features to perform the recognition. The mentioned features are the URL size, the number of hyphens, the number of dots, the number of numeric characters plus a discrete variable that correspond to the presence of an IP address in the URL and finally the similarity index. Proven by the results of this study the similarity index, the feature we introduce for the first time as input to the phishing detection systems improves the overall recognition rate by 21.8%.

Keywords: Phishing, Phishing detection system, Web security, SVM, URL, Hamming distance