Abstract

Malicious links are used as a source by the distribution channels to broadcast malware all over the Web. These links become instrumental in giving partial or full system control to the attackers. This results in victim systems, which get easily infected and, attackers can utilize systems for various cyber crimes such as stealing credentials, spamming, phishing, denial-of-service and many more such attacks. To detect such crimes systems should be fast and precise with the ability to detect new malicious content. This paper introduces various aspects associated with the URL (Uniform Resource Locator) classification process which recognizes whether the target website is a malicious or benign. The standard datasets are used for training purpose from different sources. The rising problem spamming, phishing and malware, has generated a need for reliable framework solution which can classify and further identify the malicious URL. An alternative approach has been proposed which uses a Naïve Bayes classifier for an automated classification and detection of malicious URLs. The proposed model based on Naive Bayes is supported by clustering and classification technique. On the other hand, they are rarely used for general probabilistic learning and inference which is typically used for estimating with conditional and marginal distributions. The proposed work in this paper shows that, for a wide range of benchmark datasets, Naive Bayes models learned using Probability model has better accuracy than Support Vector Machine model.
References

- Harry Zhang &quot;The Optimality of Naive Bayes&quot;. FLAIRS 2004 conference.
- DNS-BH. Malware prevention through domain blocking.
- JWSPAMSPY. E-mail spam filter for Microsoft Windows.
- PHISHTANK. Free community site for anti-phishing service.
- http://random. yahoo. com/bin/ryl3. (accessed on 20/06/2014)
- Ramchandran, A., And Feamster, N. &quot;Understanding the network-level behavior of spammers&quot;. In Sigcomm (2006).

Index Terms

Computer Science Security

Keywords

Machine Learning Feature Extraction Benign Malicious Web Pages Classification Module

Web-Based Attacks.