

‘BE LIBERAL IN WHAT YOU RECEIVE’ ON YOUR MOBILE PHONE

M. Zubair Rafique And Muddassar Farooq
{zubair.rafique, muddassar.farooq}@nexginrc.org

The volume of spam SMS received by the mobile users have dramatically increased in recent years. SMS provides a perfect model for spam and is widely exploited through arbitrary advertising campaigns and propagation of scam schemes. The increasing threat can be controlled through proficient and robust filtering systems. Filtering SMS spam is complicated by many reasons, among them is that SMS has its own specified syntax and structure and that the mobile devices are resource constrained.

In this paper we present a novel method, which incorporates the underline byte level data coding scheme of SMS to detect spam messages. Our purposed scheme is robust to word adulteration techniques and language transformations as it works on the GSM layer of mobile phone. The framework transforms the benign and spam messages features’ into Hidden Markov Models (HMM’s) through profiling of byte level distributions. This process leads to a new learning algorithm for the classification of spam SMS, which is based on the probabilistic variation from the trained models. Our framework is lightweight as it requires less processing and memory resources and easily deployable on mobile devices.

The carefully designed experiments – by accounting the rigorous test cases – results that our framework provides high detection rate and small false alarm rate in classification spam SMS. We report our experiments on a real world benign and spam datasets collected from Grumbletext and through various social communities.