Justification for Inclusion in NSPW
Position Paper: Spam Filtering Beyond Tokens

Spam is a well known problem, and filtering spam (distinguishing spam from legitimate email) has been a serious challenge. For example, keyword-based filters were quickly made obsolete by purposeful misspellings and replacement of words by images containing words. The current generation of learning filters (e.g., production-use filters such as SpamAssassin and SpamProbe that use Bayesian techniques among others) are able to handle some of the misspellings. However, there is a general problem of *adversarial learning*, where the workload can be influenced by the adversary. Consequently, there may be some inherent limitations placed on the distinguishing capabilities of token-based learning filters.

Instead of treating all tokens the same in learning filters, we advocate a semantics-based approach in this position paper. This semantics-based approach is similar to the ranking algorithms in search engines. In hub-and-authority ranking algorithms, URLs are given special meaning and treatment. By recognizing the unique value of URLs (and processing them more carefully than word tokens) in web documents, search engine ranking algorithms have been able to focus on the important documents and web sites. We want to recognize the unique value of URLs in spam messages and process the URLs more carefully than word tokens. Since most of the spam messages expect the reader to click on some URL in order to interact with the spammer, URL processing adds a reliable new dimension to token-based spam filtering. By following the links, one can gain valuable insight into the kind of websites the message is pointing to. This is particularly the case since search engines and portals (e.g., yahoo.com) have already done much of the work and can help us get around randomization and other obfuscation techniques.

We believe a careful URL processing of potential spam messages represents a new paradigm in spam filtering, the same way the addition of hub-and-authority ranking algorithms changed search engines. Due to the need to interact with the spammed victims, URLs in spam messages are typically "live," particularly during the early period when the spam has not been widely recognized by learning filters. By identifying suspicious URLs, we will be able to find a reliable indicator of spam that complements token-based learning filters. While this approach by itself is unlikely to solve the entire spam filtering problem, it seems to be a promising new paradigm in the research of spam filtering that can be combined with other spam filtering techniques to improve overall efficiency and combating spam.