

配送情報を利用した 迷惑メールのフィルタリング  
Spam filtering in use of delivery information

情報セキュリティ大学院大学  
本田 致道

<要旨>

現在、普及している迷惑メールフィルタとして、コンテンツベースフィルタとネットワークベースフィルタが挙げられる。コンテンツベースフィルタは、メールを受信した後に迷惑メールを排除するため、ネットワークやシステムへの負荷を低減させる効果は低い。ネットワークベースフィルタは、送信メールサーバの情報を利用し、迷惑メールを受信する前に排除することができるが、迷惑メールと正常メールを混在して送信するメールサーバに対して有効ではない。これらの問題に対して、Naive Bayes 分類器を用いてメールの配送情報を機械学習することによって、迷惑メールをフィルタリングする新たなネットワークベースフィルタを提案する。実際にインターネットから送信された迷惑メールに対し、提案フィルタを適用する実験を行った結果、従来のネットワークベースフィルタによる場合と比較して、フィルタを通過する迷惑メールのサイズ数を 73.4%削減する効果を得た。

<Abstract>

Currently, two types of popular spam filters; contents base filter and network base filter are used to determine if received E-mails are normal or spam. The contents base filter investigates appropriateness of words in mail message and the network base filter checks normality of sending mail servers. However, these conventional filters respectively have defects in terms of filtering performance. That is, the contents base filter is not good at reducing load on network/system resources with spam because the contents base filter processes after receiving E-mails. The network base filter, on the other hand, is not effective against the mail servers sending both spam mails and normal mails. We propose a new method of network base filtering, for solving these problems. It analyzes information for E-mail delivery, and applies Naive Bayes classifier as a machine learning algorithm. We collected actual E-mails on the Internet to compare the proposed method with the conventional methods. As a result, it is shown that the proposed method can reduce 73.4% size of spam which passed conventional filter.