ReputationBox : A System to Analyse Importance of Emails and Reputation of Email Senders

Dileepa Jayakody[#1], Gihan Dias[#2]

[#]*Department of Computer Science and Engineering, Faculty of Engineering, University of Moratuwa, Sri Lanka*

[1]dileepajayakody@gmail.com

[2]gihan@uom.lk

***Keywords*— email analysis, machine learning, email reputation**

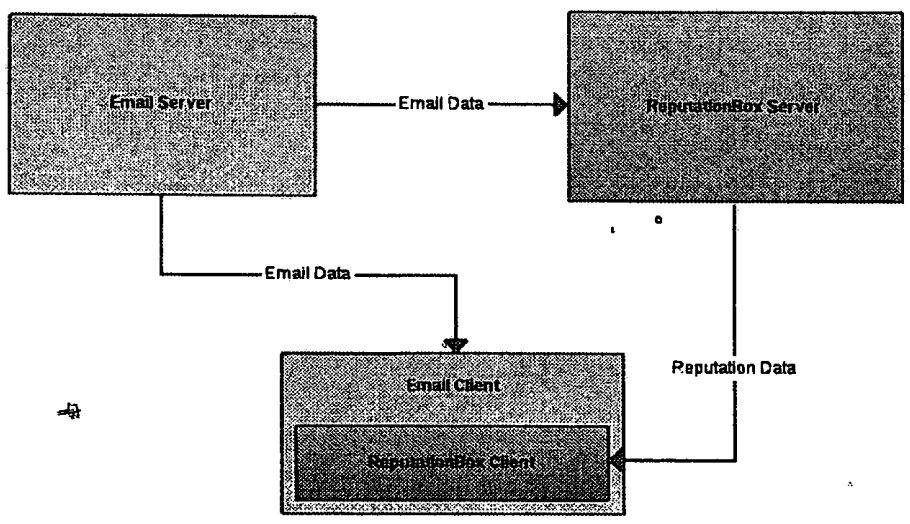
e-mail is one of the primary mediums of electronic communication used today. The number of email accounts are expected to go over 4.3 billion accounts by the year 2016 according to the latest email statistic reports. Users receive a bulk of emails on a daily basis and consequently they tend to overlook their inboxes and miss important emails from important people. This email management issue imposes an adverse effect on the productivity of email communication. Although many email clients today are equipped with tools to filter emails based on keywords, email addresses; most of these filters are static and are not updated automatically. The productivity of email communication will drastically improve if emails can be automatically evaluated for their goodness, so that users can promptly identify important and interesting emails before reading them and proceed with suitable actions on a timely manner.

The goodness of an email depends on several factors such as the relevance of email content to the user, the nature of email relationship between the user, the sender and other email correspondents addressed in the email. In this study, the measurement of email goodness is formally called as "email reputation". By looking at the reputation score of the email and the sender, user can prioritize and filter interesting and important emails effectively. With many emails and new senders accumulating on a daily basis and user's interests, priorities changing over time, it's necessary to incorporate a periodically updated email analysis model to calculate email reputation. The proposed reputation analysis algorithm uses a per-user reputation data model periodically derived from the user's past email data to evaluate email and sender reputation.

The proposed system "ReputationBox" caters for the requirement of analysing email reputation based on several features. These features include;

- People features : Nature of email relationship between the email sender, the recipient and others addressed in the email (e.g. the number of messages exchanged, average time to reply)
- Topic features : The topics and activities discussed in the email derived based on a periodically updated word space model
- Action features : Recommended user actions based on above people and topic features and correlated user actions with past email data (e.g.:seen, read, replied, deleted emails including certain topics and email correspondents)

FIGURE I
REPUTATIONBOX SYSTEM ARCHITECTURE



ReputationBox system has 2 main components.

1. ReputationBox server: Performs the main analysis process of email reputation
   - Email Connector: Connects to the user's mailbox, retrieves emails periodically
   - Reputation Analyser: Performs email analysis, computes email and sender reputation and recommends suitable action for new emails
   - ReputationStore: Stores the reputation meta-data for each email and email correspondent
2. ReputationBox client: Installed in an email client as a plugin. When a new email comes to the inbox, it retrieves reputation data from the ReputationBox server and displays the reputation profile and recommended action for the corresponding email

The ReputationBox server periodically retrieves user's emails and triggers the email analysis process to update the reputation data model. Subsequent to the email analysis process, the server persists the reputation data so the ReputationBox client can retrieve the reputation results over a RESTful web service and display the results with the corresponding email. By connecting the mailbox with the ReputationBox system, users can manage their emails effectively based on email reputation results and recommended email actions.