New Bio Inspired Techniques in the Filtering of Spam: Synthesis and Comparative Study

Hadj Ahmed Bouarara, GeCode Laboratory, Department of Computer Science, Dr. Moulay Tahar University of Saïda, Saïda, Algeria
Reda Mohamed Hamou, GeCode Laboratory, Department of Computer Science, Dr. Moulay Tahar University of Saïda, Saïda, Algeria
Abdelmalek Amine, GeCode Laboratory, Department of Computer Science, Dr. Moulay Tahar University of Saïda, Saïda, Algeria

ABSTRACT
The internet era promotes electronic commerce and facilitates access to many services. In today’s digital society the explosion in communication has revolutionized the field of electronic communication. Unfortunately, this technology has become incontestably the original source of malicious activities, especially the plague called undesirables email (SPAM) that has grown tremendously in the last few years. This paper deals on the unveiling of fresh bio-inspired techniques (artificial social cockroaches (ASC), artificial haemostasis system (AHS) and artificial heart lungs system (AHLS)) and their application for SPAM detection. For the authors’ experimentation, they have used the benchmark SMS Spam corpus V.0.1 and the validation measures (recall, precision, f-measure, entropy, accuracy and error). They have optimising the sensitive parameters of each algorithm (text representation technique, distance measure, weightings, and threshold). The results are positive compared to the result of artificial social bees and machine learning algorithms (decision tree C4.5 and K-means).

KEYWORDS
Bio-Inspired, Haemostasis, Heart Lungs System, Social Cockroaches, Spam Detection, Validation Measure, Visualisation

1. INTRODUCTION
In the 1990s, the Internet has become a consumer tool that plays an important role in the functioning of our society especially with the development of the web services such as email that is used daily by millions of persons. Unfortunately, this technology provides a real problem for individuals and companies due to the increasingly huge amounts of junk mail (spam). The messages are classified as spam if they meet the three following criterions:

- They have not been solicited and have no interest.
- They are sending in bulk.
- The user does not know the sender

Today, spam is one of the most annoying things in the web. According to the statistics provided by the Radicati group in 2014 who supplies quantitative and qualitative researches with details on the e-mail, the security, the instant messaging (IM) and the social networks there was exactly:

DOI: 10.4018/JITR.2016040103

Copyright © 2016, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
• 3 million of spam messages were sent per second.
• 262 billion of spam emails were sent per day.
• 95,000 billion spam emails were sent in this year.
• 75% of SMTP traffic was generated by spam on the Internet.

This phenomenon represents a big nuisance to the traffic because some of them can be very dangerous for your computer (viruses, worms, Trojan -horse) as for yourself or your children (for example pornographic website). This form of email must be immediately deleted.

The tremendous increase of unsolicited mail (spam) over the past years (Geer, 2004) has rendered email communication without spam filtering almost impossible. In the last few years a range of filtering solutions have emerged with a role is to detect unwelcome emails, some of them are based on the identification of the path travelled by the messages and others are based on the content of messages. It is a sophisticated and challenging problem as long as spammers continue developing new methods and techniques that are being used in their campaigns to defeat and confuse email spam filtering process. Furthermore, the majority of the classical spams filtering systems suffer from several drawbacks in terms of precision (classification of unwanted message in the inbox and classification of legitimate mail as a spam), response time and presentation of the results.

To deal with the previous drawbacks a new paradigm has emerged named bio-inspired and several filtering systems were developed inspired from the biological life of living systems such as insect lifestyle or from the most known natural phenomenon. This paper is positioned in the framework of detailing a set of ours new bio-inspired techniques (artificial heart lungs system (AHLS), Artificial social roaches (ASR), artificial coagulation system (ACS), social worker bees (SWBs)) and theirs application for the spam filtering problem in the form of a synthesis and comparative study.

1.1. Aims and Answers

• Detailing the functioning of the algorithms (artificial social roaches (ASR), social worker bees (SWBs), artificial coagulation system (ACS) and artificial heart lungs system (AHLS)) and how they are applied to the problem of spam filtering.
• Analysing the different results obtained after applying the techniques (AWLS, ASR, ACS, SWBs) on the dataset spam V.0.1.
• We will answer the following questions that can help others researches in future:
  • Why we have chosen the biological coagulation system, the biological heart lungs system, the lifestyle of social roaches and the lifestyle of social workers bees to elaborate a new bio inspired spam filtering system?
  • What is the ideal configuration of each algorithm?
  • Which technique is the most valuable?
  • Does the bio-inspired algorithm (ASR, ACS, SWB and AHL) ensure more security and protection against the SPAM compared to the machine learning techniques?
• Summarizing the functioning of the different algorithms in a table?
• What is the text vectoring technique that ensures the best coding of the emails?

The paper is organized as follow: Section 2 discusses some classical anti-spam systems existed in literature; section 3 detailed the functioning of ours new bio-inspired spam filtering techniques (artificial heart lungs system (AHLS), Artificial social roaches (ASR), artificial coagulation system (ACS), social worker bees (SWBs)); section 4 regroups: i) the results of the bio-inspired spam filtering techniques (AHLS, ASR, ACS and SWBS) ii) ideal configuration of each algorithm ; section 5 comparing the results of the bio-inspired spam filtering systems with the results of the machine learning spam filtering techniques (K-means and decision tree C4.5). Finally, section 6 concludes the paper and gives some future works.
A Case Study of One IT Regional Library Consortium
[www.igi-global.com/chapter/case-study-one-regional-library/44517?camid=4v1a](www.igi-global.com/chapter/case-study-one-regional-library/44517?camid=4v1a)

Humanitarian Logistics: A Review and Scientometric Analysis
[www.igi-global.com/article/humanitarian-logistics/212609?camid=4v1a](www.igi-global.com/article/humanitarian-logistics/212609?camid=4v1a)