Optimization of Anti-Spam Systems with Multiobjective Evolutionary Algorithms

Vitor Basto-Fernandes, School of Technology and Management, Computer Science and Communication Research Centre, Polytechnic Institute of Leiria, Leiria, Portugal

Iryna Yevseyeva, School of Technology and Management, Computer Science and Communication Research Centre, Polytechnic Institute of Leiria, Leiria, Portugal & TEMA - Centre for Mechanical Technology and Automation, University of Aveiro, Aveiro, Portugal

José R. Méndez, University of Vigo, Ourense, Spain

ABSTRACT

In this paper anti-spam filtering is presented as a cumbersome service, as opposed to a software product perspective. The huge human effort for setting up, adaptation, maintenance, and tuning of filters for spam detection in anti-spam systems is explained. Choosing the best importance scores for the spam filters is essential for the accuracy of any rules based anti-spam system, and is also one of the biggest challenges in this research area. Optimal filters score settings for Apache SpamAssassin project (the most widely adopted anti-spam open-source software) is addressed. In addition to a survey done on single/multi-objective optimization research in this area, we also present a study for filters score setting using multiobjective optimization based on two most representative evolutionary algorithms, NSGA II and SPEA2. Problem description, simulation and results analysis is done for SpamAssassin public mail corpus which is widely used for benchmarking purposes.

Keywords: *Anti-Spam, Evolutionary Algorithms, Multi-Objective, Optimization, SpamAssassin*

INTRODUCTION

E-mail and Web applications were responsible for the massive adoption of the Internet for personal, business and governmental usage in the last two decades. Malicious usage of electronic data distribution and all other forms of unsolicited communications, also designated as spam, has reached scales never seen before.

Every day e-mail users receive lots of messages containing unsolicited, unwanted, legal and illegal offers for commercial products, drugs, fake investments, etc. Spam traffic has increased exponentially in the last few years. During September 2010 the percentage of spam deliveries accounted for about 92% of all Internet e-mail traffic (MessageLabs Ltd., n.d.). The number of messages arriving to a mail server can easily reach the order of a million per month for small organizations or be in the order of a million per

DOI: 10.4018/irmj.2013010105

day for a medium/big organization. Estimates on worldwide cost of spam in each of the last few years are of hundreds of billions U.S. dollars (Schryen, 2007), mainly due to loss of productivity for users and costs of setting up and maintaining anti-spam systems.

Although e-mail has represented the main distribution channel of spam contents due to its low cost and fast delivery characteristics, Web became recently also a target for spam distribution. The change of the strict publishing-consumer approach of Web 1.0 to the collaborative approach of web 2.0, adopted by Content Management Systems (CMS), where every user is able and stimulated to produce, publish and share data, made it attractive for spam to be spread through Weblog posts, Wikis, social networks, virtual communities, etc., in addition to mobile Short Messaging System (SMS) advertising.

The traditional e-mail services have been modified, with varying degrees of success, to adapt to this type of attacks that are able to block e-mail servers completely. The cost of transmitted messages bandwidth, processing time, storage and especially time spent by users to manually identify and remove spam messages is alarmingly high (reaching several days a year devoted to spam sorting (Schryen, 2007) and follows the trend of spam traffic growth. The problem becomes critical in recently fast growing communities of mobile device users (e.g., Android, Blackberry, etc.), mainly because of mobile devices considerably reduced resources.

Current solutions for filtering spam are often based on centralized or distributed trusted and untrusted servers lists. There are also solutions for message content analysis, but these apply only to a limited scope (only text, neither images nor PDFs). They introduce probabilistic uncertainty in the processing of mail and require a comprehensive maintenance for the filters to properly identify the types of messages that must be accepted or not. Methods of sending spam are continuously refined and adapted to most common and up to date filters, forcing anti-spam system administrators to constantly react and upgrade their system in a permanent race against spammers.

Several hundreds of complex filters are used in initial distributions of anti-spam systems and more filters are added in a regular basis. Importance and tuning of each of them depends on system, type of organization, business domain and requires heavy manual configuration and maintenance. Anti-spam filters are also context (location, language, culture) dependent and anti-spam tools based on the analysis of messages need to be tuned to local, specific contexts. Most popular and general anti-spam tools are optimized primarily for the spam in United States of America, being not so effective for spam filtering messages in other languages.

Anti-spam systems aim for manual work reduction on spam-filters tuning, configuration, maintenance and filters adaptation to the context or operation domain. Due to the very high amount of messages to be classified in very short time by anti-spam systems, high performance algorithms for filters processing are needed in order to minimize classification processing time.

Spam Filtering Approaches

Due to the high complexity of spam classification, current solutions are based on the combination of multiple techniques of different types, namely collaborative, content-based and domain authentication techniques.

Collaborative filtering is based on the usage of different protocols and tools, which allow exchanging information on spam messages and source servers (spam e-mails or servers that have been used to distribute spam). The flexibility of the Domain Name System (DNS) protocol allows sharing information on whether a server is a source of spam or not. This gave rise to two well know techniques named white lists (i.e., Dnswl.org, 2007) and black lists (i.e., SpamHaus lists; SpamHaus Project Organization, n.d.), DNSWL and DNSBL, respectively. Peer-to-peer (P2P) complex systems were also created to share spam messages signatures