E-Mail System for Automatic Hoax Recognition

Tomislav Petković, Zvonko Kostanjčar and Predrag Pale
Department of Electronic Systems and Information Processing
Faculty of Electrical Engineering and Computing, University of Zagreb
Unska 3, 10000 Zagreb, Hrvatska
Email: tpetko@zesoi.fer.hr

***Abstract*— With the advent of Information society false and inaccurate information represents one of the major problems. Hoaxes and unsolicited commercial e-mail messages (SPAM) are an important example of such information. A conceptual solution together with the developed system for automatic hoax recognition is presented. Hoax recognition is done in several parallel steps to increase system accuracy and robustness. Each step is implemented as a separate module that outputs measure of similarity. Fuzzy logic expert system makes the final decision whether the received mail is an actual hoax. The system is available as a free e-mail service of the Croatian CERT**

I. Introduction

Rapid growth of e-mail at rates up to 66% annually can not be attributed to increase in person-to-person communication only [1]. Major part should be contributed to spam and automatically generated messages such as virus alerts and other notifications. Only some smaller part of increased e-mail traffic can be attributed to hoaxes.

In a modern Information Society false and inaccurate information represents one of the major problems. Unsolicited commercial e-mail messages (SPAM) are an example of mostly inaccurate and unwanted information, and sometimes even completely false information with hoaxes being the major example. Junk e-mail and spam are major problem as each user wastes time sorting the unwanted e-mail, but hoax e-mail messages present another problem for all Internet users because in some cases even more knowledgeable users are not able to discern true from false information.

A. Hoaxes and spam

Hoax is an email message containing bogus contents with intention to mislead or even scare the person receiving it. The goal behind it is to have the message forwarded to as many recipients as possible [2]. Most common forms of a hoax according to the CARNet CERT are fake virus warnings, chain mails, false help requests, threatening and/or scaring messages, false petitions, compromising hoax and harmless hoaxes [2].

This definition of a hoax should be expanded to stress the importance of intention to mislead the average user. Unsolicited commercial e-mails also frequently contain bogus contents and misleading information, but such content is usually composed of overstatements, exaggerations and omitted information with the sole purpose to sell the product. In the case of unsolicited commercial e-mail messages the average e-mail user is able to recognize the message as the advertisement and consequently will expect that the information contained within is not entirely true or is incomplete. Hoax is usually crafted in such way that the average user is not able to recognize it as false and is compelled to act in accordance to the content of the hoax.

So hoaxes should be described as a messages transmitting false information with malicious intention to mislead or scare the person receiving it. The goal behind is to spread the false information and to make the recipient act in accordance to the content of the hoax.

II. Hoax Recognizer

Many various solutions exist for automated text classification or, more specifically, for automated e-mail classification [3], [4]. Most are concerned with message management [3] and some are specifically crafted for spam and junk e-mail filtering [5], [6].

Hoax Recognizer system is primarily designed as a web service for hoax recognition. The user sends e-mail message believed to be a hoax to the system and receives the results of the recognition. The actual process is shown on figure 1.

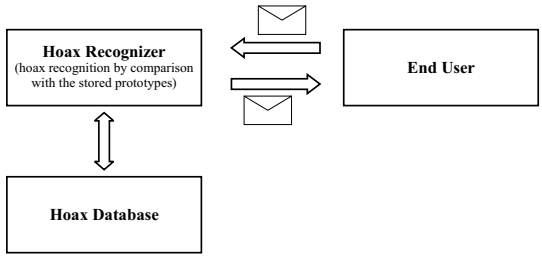

Fig. 1. Hoax Recognizer System Overview

A. System overview

When the Hoax Recognizer system receives the message from the user recognition process begins. In a proposed system recognition is done in several parallel steps to increase system accuracy and robustness. Each step is implemented as a separate program module. Output of each module is a measure of similarity defined as a real number from the zero to one interval, where one denotes a certain hit (a hoax). Fuzzy logic expert system makes the final decision whether the received mail is an actual hoax.

It is important to note that system compares received message with the stored hoaxes and hoax prototypes. Because of that most significant limitation of the system is inability to recognize new hoaxes with novel concepts—to enable the system to reliably recognize new hoaxes database update is usually needed.

B. System modules

Implemented system has four comparison modules [7] that perform the comparison independently. First two modules are based on simple distance measures between two text messages. Modified nearest neighbor algorithm is used for text classification with the edit or Levenshtein metric used for distance calculations [8], [9]. Edit or Levenshtein distance between two strings is defined as minimal number of edit operations to make two strings equal. Edit operations consist of insertions, deletions or changes of character in a string.

First module calculates edit distance between received e-mail message and known hoaxes stored in a database. Depending on the calculated distances module classifies received e-mail as one of the known hoaxes or as an unknown message. The prototype for regular e-mail messages does not exist and examined message is classified as a regular mail if the distance from all the hoax prototypes is large enough.

Stored hoax prototypes used for distance calculation in second module contain characteristic phrases, sentences and paragraphs as this module extends edit distance to sentences and paragraphs. Edit distance between sentences is defined as minimum number of word insertions, deletions or changes needed to make two sentences equal. Furthermore, different weights to insertion, deletion or changes are added based on word length. Distance between two paragraphs is similarly defined. For each received e-mail message distances for words, sentences and paragraphs are calculated. Because edit distance is used exact word ordering in a sentence or sentence placement in a paragraph does not represent a problem when recognizing a hoax. As additional benefit, phrases comparison can detect changed hoaxes or chain letters thus detecting new false messages that were created by some malicious user.

Edit or Levensthein distance satisfies the conditions: $\forall \vec{x}, \vec{y} \in \mathbf{X} : \; d(\vec{x}, \vec{y}) \geq 0$ with $d(\vec{x}, \vec{y}) = 0 \Rightarrow \vec{x} = \vec{y}$, $d(\vec{x}, \vec{y}) = d(\vec{y}, \vec{x})$ (symmetry) and $\forall \vec{x}, \vec{y}, \vec{z} \in \mathbf{X} : \; d(\vec{x}, \vec{y}) \leq d(\vec{x}, \vec{z}) + d(\vec{z}, \vec{y})$. Furthermore, edit distance is bounded. If we define $|\vec{x}|$ as the number of letters in a word $\vec{x}$ and assign the equal value to all of the operations (insertion, deletion and change)

$$0 \leq d(\vec{x}, \vec{y}) \leq \max\big(|\vec{x}|, |\vec{y}|\big) \tag{1}$$

holds for any two words $\vec{x}$ and $\vec{y}$. This must be taken into consideration when calculating similarity measure between two different texts. With the edit distance between two sentences upper bound becomes the number of words in a longer sentence. Same applies for paragraphs in a text. When defining the hoax prototypes for the nearest neighbor algorithm the length of the prototype is known beforehand, but the length of the message to be compared is not. When calculating the distance the upper bound is $\max\big(|\vec{x}|, |\vec{y}|\big)$. If we take the length of the prototype as the maximum distances between the prototype and any other message if that incoming message is shorter no error is introduced. In the case when the incoming message is larger then the prototype and if the calculated distance is greater then the length of the prototype we can assume that the incoming message is completely different and is not a hoax. This is justified as the distance is symmetrical and result larger than the length of the prototype can be interpreted as the necessity to completely change the prototype. The neighborhood of the observed prototype should include everything nearer then 2 or 3 edit operations to allow most common spelling errors and changing the word order in a sentences.

As the maximum distance is known measure of similarity can be calculated easily. Edit distance between words and edit distance between sentences is calculated. Let d_1 be edit distance between words and d_2 distance between sentences. Furthermore, let us denote with m_1 maximum distance between words and with m_2 maximum distance between sentences. Measure of similarity can then be defined as

$$\text{MS} = 1 - c_1 \frac{d_1}{m_1} - c_2 \frac{d_2}{m_2} \tag{2}$$

with the $c_1 + c_2 = 1$. For Hoax Recognizer values of 0.75 for constant c_1 and 0.25 for c_2 were chosen thus giving greater importance to distance between words.

Even simple hoax examination reveals that vast majority of such messages contains at least some characteristic phrases. Third module utilizes statistical methods primarily for phrase comparison between potential hoax message and stored prototypes. For all characteristic phrases a table [10] which states the number of appearances in two texts under comparison is computed.

Modules described so far are primarily focused on message content and structure (i.e. text organization in words and sentences). Fourth module attempts to capture the meaning of the message. If two text messages are conveying same information it is correct to assume existence of a set of mutual words. We observe quantity of new information contained in potential hoax message. If such observed message does not contain new information compared to the stored hoax prototype two messages can be considered the same. Obviously, if new information exists messages are different.

Additional project documentation which describes actual implementation of the described subsystem and the entire system is available online on the Internet [7], [2].

C. Example runs

As stated before the end user must send the message believed to be potential hoax to the Hoax Recognizer system (figure 1) and will receive recognition results by e-mail. Two sample system runs for a simple message and for actual Hoax message are shown on table I.

There exists a possibility that received message is a new hoax but it will not be recognized. As each separate subsystem gives measure of similarity computed measures can be used to decide when to alert the system administrator. If the measure of similarity falls in the area between sure recognition of a hoax and sure recognition of a normal message administrator will be alerted. Administrator then must check whether received message is a hoax and act accordingly. Described situation is show on figure 2. If administrator determines that the received message is new type of a hoax system database must be updated. In that

TABLE I
EXAMPLE RUNS

INCOMING E-MAIL MESSAGE IS NOT A HOAX

Example Query		Example Response (in Croatian)	
To:	hoax@cert.hr	To:	tomislav.petkovic@inet.hr
From:	tomislav.petkovic@inet.hr	From:	hoax@cert.hr
Subject:	Test	Subject:	Odgovor na vasu prijavu
Ova poruka sigurno nije hoax. This message is not a hoax. – tomislav.petkovic@inet.hr		Postovani korisnice, Ovo je automatski generirana poruka. E-mail koji ste nam proslijedili ne lici na neki od poznatih hoaxa. Detaljnije cemo analizirati Vasu poruku i obavijestiti Vas radi li se ipak o hoaxu. Pozdrav, CARNet CERT www.cert.hr		

INCOMING E-MAIL MESSAGE IS A HOAX (MONEY TRANSFER SCAM)

Example Query		Example Response (in Croatian)	
To:	hoax@cert.hr	To:	tomislav.petkovic@inet.hr
From:	tomislav.petkovic@inet.hr	From:	hoax@cert.hr
Subject:	Test	Subject:	Odgovor na vasu prijavu
DEAR TOMISLAV, My name is MUSTAPHA NDOH, a citizen of BOUAKI REPUBLIC OF COTE D IVOIRE. I got your contact from the Chamber of Commerce here during my search for an international business relationship. I am residing in Abidjan the capital city of Cote dIvoire. In fact I worked with the ROBERT GUEI as one of his personal confidance.During the renewed political crisis of which the he was brutally mudered by the government forces who claimed it to be the rebels who has been causing problems in my country. It was then that I moved the total sum of US $25,000,000.00 (Twenty Five Million United States Dollars) which was from the sales of cofee and cocoa and gold…		Postovani korisnice, Ovo je automatski generirana poruka. E-mail koji ste nam proslijedili prepoznat je gotovo sigurno kao hoax ”Nigerian Scam-en-ver8”. Za vise informacija, pogledajte na http://www.cert.hr/hoax.html A sadrzaj uzorka po kojem je prepoznat ovaj hoax mozete pogledati na: http://www.cert.hr/hoax_indetail.php?hid=3632 Nadamo se da Vam ovaj Hoax nije prouzrocio bilo kakve neugodnosti. Molimo Vas da i ubuduce bilo kakve sumnjive poruke dobijene elektronickom postom proslijedite nama. Pozdrav, CARNet CERT www.cert.hr		

case additional message will be sent to the user who reported the incident.

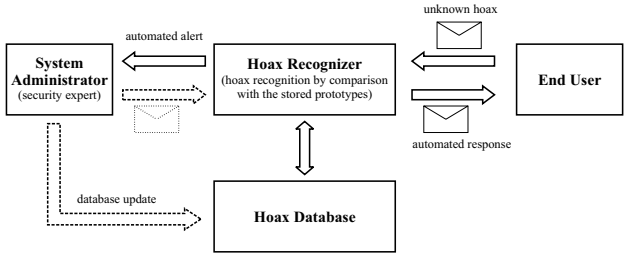

Fig. 2. Unknown or unrecognized e-mail message believed to be a potential hoax

III. RESULTS

The system is available as a free e-mail service of the Croatian CERT (http://www.cert.hr/). Any internet user can send a message believed to be a potential hoax to hoax@cert.hr and will immediately receive recognition results by e-mail. System has been operational for past year and extensive usage data has been collected. This data represents a reliable description of hoaxes and spam messages in Croatian cyberspace.

For actual implementation core of the system is implemented in C programming language. Database management system is provided by MySQL and Procmail is used to process incoming messages. Although core of the system was implemented in C calculating edit distances for words and sentences is a heavy computational task and could not be done in real-time. Beacause of such heavy computational load Hoax Recognizer system was implemented as a offline e-mail system.

A. Hoaxes in Croatia

As opposed to the rest of the world in Croatia hoaxes did not have significant share in total e-mail traffic for the past few years. This can be attributed to the small number of internet users.

In the period from the 19th may to the 27th september total of 371 e-mail messages was received (table II). Of those 273 were real queries and the rest can be classified as spam and comerical junk mail. From those 273 messages 74 messages were recognized as hoaxes (reported measure of similarity is greater than 0.5). Most of the disclosed

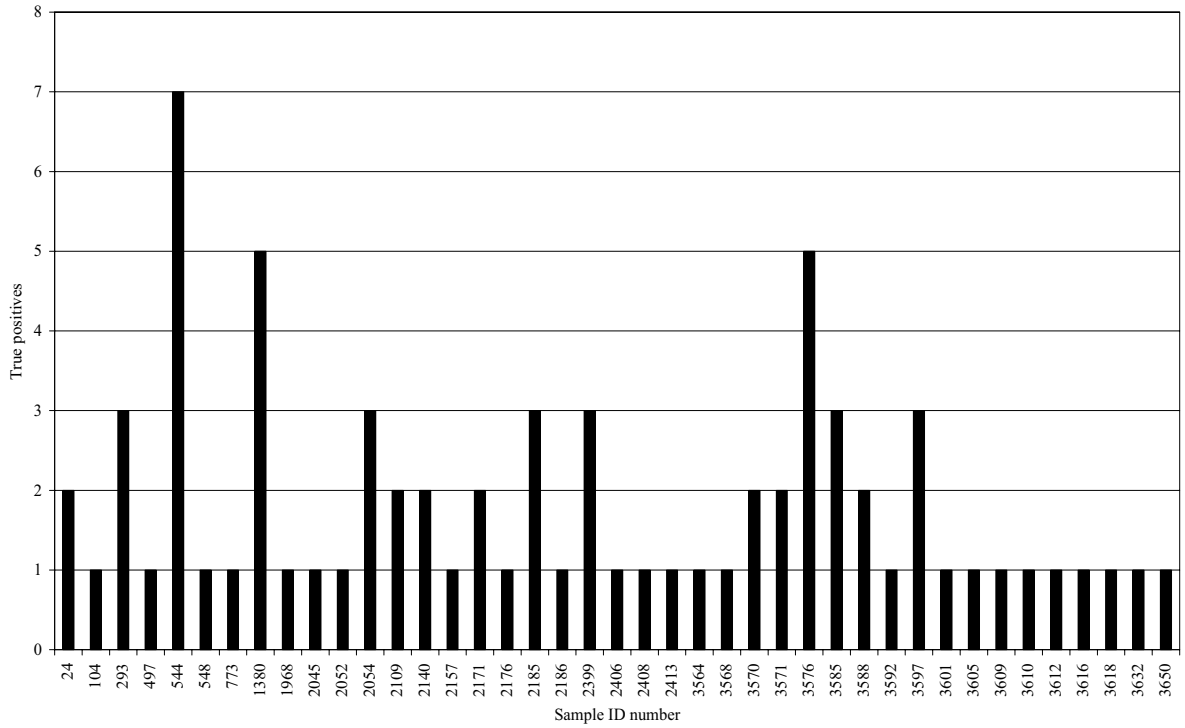

Fig. 3. Recognition results for the period from may to september 2004

hoaxes in Croatian cyberspace were written in English, the unofficial universal communication language in Internet. The rest of the messages (199 in total) were short test messages[1]. Analysis of the user structure shows that the majority of users are CERT personel—of 371 total queris they have used the service 134 times.

Histogram of occurences for 74 received e-mail messages classified as a hoax is shown on figure 3. Most of the received hoaxes appear only one or two times (average is 1.80). If we compare most occuring hoaxes by content almost all of the more frequent ones are some variant of money transfer scam (table III).

TABLE II
MESSAGE STATISTICS (MAY TO SEPTEMBER)

E-mail type	Number of occurences
Spam	98
Hoaxes	74
Other messages	199
Total number of received messages	371

TABLE III
MESSAGE STATISTICS (TRUE POSITIVES)

ID	Sample name	Number of occurences
544	Money Transfer hoax-us-ver3	7
1380	FROM THE DESK OF EDMOND ZAMFARA Hoax-us-ver1	5
3576	lotto scam-en-ver3	5
293	Foreign account scam Hoax-us-ver3	3
2054	zuma hoax-us-ver1	3
2185	mr biko hoax-us-ver1	3
2399	Famous lottery Hoax-us-ver1	3
3585	Money transaction-en-ver2	3
3597	Mobile Phone Virus Hoax-en-ver1	3

IV. CONCLUSION

In this paper we have presented an automated system for hoax recognition. As the recogition step utilizes edit distance presented Hoax Recognizer system is immune to most spelling errors and perturbation of words in a sentence that are common in electronic messages. However, main disadvantage of the presented system is computational time required to classify single message. Because of the heavy computational requirements system was implemented as an offline e-mail system.

Hoax Recognizer has been operational for past two years as the public web service of the Croatians CARNet CERT. From the collected data detailed analysis of the hoaxes in Croatian cyberspace can be made. However, as the major users of the system are CERT personel collected data mainly reflects reported incidents.

[1]Vast majority of such messages contained only word test or one or two simple sentences.

ACKNOWLEDGMENT

The authors would like to thank CARNet CERT and LS&S for given support. Financial support was provided by CARNet CERT, contract number 0114-24/147-2001.

REFERENCES

[1] S. Hinde, “Spam, scams, chains, hoaxes and other junk mail—discussion,” *Computer & Security*, vol. 21, no. 7, pp. 592–606, Nov. 2002.

[2] (2003) CARNet CERT Hoax Recognizer website. hoax.html. [Online]. Available: http://www.cert.hr/

[3] E. Crawford, J. Kay, and E. McCreath, "Automatic induction of rules for e-mail classification," in *Proceedings of the Sixth Australasian Document Computing Symposium*, Coff Harbour, Australia, Dec. 2001.
[4] C. Rudolfs, "E-sl@ve, an incremental approach to automated, content-based email classification," Master's thesis, Department of Computing Science, Katholike Universitat Nijmegen, Nijmegen, The Netherland, Aug. 2002.
[5] S. Hird, "Technical solutions for controlling spam," in *Proceedings of AUUG2002*, Melbourne, Australia, Sept. 2002.
[6] (2004) The Apache SpamAssassin project. [Online]. Available: http://spamassasin.apache.org/
[7] (2002, Feb.) Hoax Recognizer Documentation v0.1.3. index.html. [Online]. Available: http://ipg.zesoi.fer.hr/petkovic/hoax_recognizer/
[8] G. Navarro, "Approximate text searching," Ph.D. dissertation, Department of Computer Science, University of Chile, Santiago, Chile, Dec. 1998.
[9] V. Levensthein, "Binary codes capable of correcting spurious insertions and deletions of ones," *Problems of Information Transmission*, vol. 1, no. 1, pp. 8–17, Jan.-Mar. 1965.
[10] Ž. Pauše, *Uvod u matematičku statistiku*. Zagreb, Croatia: Školska knjiga, 1993.
[11] K. Sikić, Z. Kostanjčar, N. Mareković, and T. Petković, *Prepoznavanje hoaxa*, Laboratorij za sustave i signale, Unska 3, Mar. 2002.