SECURITY ISSUES OF EMAIL MARKETING SERVICE

Liberios VOKOROKOS, Ján HURTUK, Branislav MADOŠ, Peter OBEŠTER
Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Letná 9, 042 00 Košice, Slovak Republic, tel. +421 55 602 3023,
e-mail: {liberios.vokorokos, jan.hurtuk, branislav.mados}@tuke.sk, peter.obester@student.tuke.sk

ABSTRACT

The paper deals in its first part with the topic of email marketing as well as with the topic of sending undesired messages when performing email marketing. The paper discusses industry standards and laws according which the email messages are to be compliant. Special attention is paid to the unwanted emails detection. Concept of solving the issues of spreading undesired emails when providing the service of email marketing is also described.

Keywords: *email marketing, undesired emails, spam,*

1. INTRODUCTION

Composing and writing email messages or even responding to those messages are the most common duties people experience during their working hours every day. Since people come into contact with email and social network every day, there is a space for using these phenomena to promote products and services. Email marketing along with social networks marketing represent a modern, effective and widely spread way of promoting services and products using email service. Taking into account the fact that in these days most people have an email address or social network account, this type of marketing is quick, simple and effective way of providing customers with news. The mentioned way of promoting business is used greatly mainly in abroad countries. Foreign companies use it to widen and, at the same time, keep their base of customers via this permanent contact. Based on these facts, the companies create databases of their clients and spread their influence in the market.

2. E-MAIL MARKETING

According to statistics, email marketing and social network marketing are one of the most effective marketing, promoting and communication tools. The main advantages of email marketing as a business and communication tool are listed below:

- delivery of an email message is requested from customers themselves, respectively receivers,
- customizing campaign according to desired criteria,
- quick carrying out of campaigns,
- low initial costs to start and promote campaigns,
- relatively high investment returns,
- active communication with customers,
- spreading campaign via customers, so called viral marketing,
- measuring and evaluating the statistics of campaign`s effectivity,
- using automated answers to its activation.

Listed advantages are but a few most important features that can be found when dealing with email marketing. One of the greatest pros of this kind of marketing is a fact that unlike passive marketing websites, this one is rather pro-active. This means that the difference is made by the user spreading news actively and not waiting for the interaction from the customer side. It is worthy to mention the personalization of the messages which leads to a more effective campaign. Regarding sending aimed advertisement in the form of email messages, it is crucial to administrate following rules and practices:

1. Regular sending – it is a good practice to keep regular promoting messages sending. It is crucial, though to set the time period between message deliveries to a comfort span so as not to disturb the customer.
2. Information value and its recency – information has got to be up to date and its content should have informative value. From customer`s perspective, it is better to receive fewer valuable messages than lots of emails without much to offer.
3. Ratio between information value and business information – this rule dictates to send a bigger amount of information interesting to the customer in favour of fewer business information.

Apart from those criteria, each and every marketing email message should be compliant with the RFC 5322 standards, which advices the message to meet the following requirements:

- message header – contains information about message routing, i.e. sender address, recipient address, message subject, date, message type etc.
- message body – contains the message itself, which can be delivered in following formats:
  - HTML - html format is produced in case the email client of the recipient is able to interpret HTML tags. Such a body contains formatted message text along with hyperlinks.
  - Plain text – this is used when email client is not able to interpret HTML tags, thus message is displayed as the text.

    - HTML and plain text format – in case of sending an email to a large number of customers it is advised to insert both versions of the same message and the client on the recipient side decides which format to render.
- envelope – contains current routing information, i.e. communication between client and email server during SMTP communication.

Provided the email marketing system is set up incorrectly and the already mentioned criteria are not met, this activity may seem like sending spam mail. Although email servers contain filters to filter spam mail, the recipient cannot always be protected from the junk mail.

Based on duties and rules dictated by laws, it is important to keep to following points when running email marketing:

1. Voluntary subscription to receiving email messages - it is a good custom to let recipient choose to voluntarily receive emails to their mail addresses. This is usually carried out twice, first time when a form is filled out, second one when the user is asked to click on a link sent to him via mail.
2. Signing out from subscription – each email message should have a link to let recipient end their subscription from marketing campaign. After this is successfully done, the user should have be kept with a chance to subscribe again in case they change their mind.
3. Not to use bought email addresses - using bought databases of email addresses decreases the effectiveness of a campaign immensely. This is due to several reasons. Firstly, the database may contain email address of people not agreeing to be a part of this campaign, which makes them undesired part of it, for they do not have any interest in offered services. Such a recipient may mark the mail as undesired and decrease the trustworthiness of campaign. Secondly, databases like these may involve addresses that are not up to date, not valid or even fake. These facts also make campaign and its creator to be less trustworthy [1].

3. UNDESIRED E-MAIL

Within this category, all unwanted or malicious sent messages with the same content are meant. In most cases, unwanted email is a form of a forced means of advertisement and can be classified as misuse of electronic communication.

3.1. Unwanted mail qualification

Main reason for sending unwanted mail is to notify recipient of services or products to gain profit, regardless the recipient, i.e. mail box owner interest in such a service or product. Another common reason for spreading such mail is trying to gather personal information about recipients and further misuse of the gathered data or their selling to third parties. In the area of the Internet, 2 sorts of spam are known:

1. overloading discussing groups – this kind of spam is aimed on discussing groups (forums) and social networks. It is based on adding feeds to the groups in order to promote products or pornographic material. These feeds are commonly very different from the topic of a specific group or social network. It is mainly focused on people with lesser IT knowledge who are tempted to visit every web site promoted.
2. email spam – this type of spam is based on sending unwanted email to specific users on their personal accounts. Attackers (spammers) gain email address from discussing groups or programmed boots called crawlers. Their job is to search through web sites to download the addresses. Another way is providing an on-line questionnaire. Users fill that in along with their addresses – unknowingly giving the creator of such a form their mail boxes. Addresses gathered like this are sold to the third party.

Email spam can be considered as a load of sent messages with unwanted commercial content. Currently, 90 per cent of all email communication world-wide can be considered as spam. From the financial perspective, email is extremely cheap and thus to senders benefits. Based on these facts, spam sending became automatic, resulting in systems to send junk mail.

Email spam according to its content can be separated as follows:

1. business offers linking to web sites – type of spam considered to be most famous and spread. Many products and services are promoted in this way, including fake products of originals, financial games, insurance offers or selling every day products. Apart from this, linked web site can be used to gather personal information about people. This kind of gathering information is called phishing. This represents a fake of a legal web site which is unrecognizable from its original designed to mislead the user to get login credentials (for instance, to get access to internet banking).
2. „Nigerian letters“ – a group of fake email messages informing recipient about their winnings in a game or gaining profit from imaginary business operation. In most cases, conditions upon which the sender asks personal information are listed, or even payment in advance is required. Based on this we may state it is clearly illegal.
3. Contact emails – this group of emails consists of messages sent in loads and are formulated in ways their recipients are asked to reply to them. The sender receives a checked list of email addresses which is later misused for spreading spam.
4. Verification emails – they are used to verify email address of the receivers. This kind of message includes a link to website with encoded

information about a recipient. Upon clicking on the link or downloading verification image, the address is automatically stored on the side of the sender and such an address can be used in campaign.

5. viruses – these email messages include links to websites with malicious code or such code is already a part of the received message in form of javascript or .exe file.

Email spam spreading can be undertaken in following ways:

1. Direct spam sending from its creator – this way is not very popular, for the creator can be very quickly found according to the IP address from which the spam is sent and the address is written to blacklist. Based on the list, it is possible to block this IP address on the side of email server.
2. Sending spam from computers infected by malicious code – this way is getting more and more popular. In this case, it is needed to infect computers with malicious code – bot, or even a group of infected computers called BOTNET. Individual bots are controlled by main initializer and they are given information about recipients to whom they send messages. This is a very effective way, as one bot can send hundreds of messages in a single minute.

Both ways of sending undesired mails benefit from poorly secured ISP servers, commonly known as open-mail-relay. Those ISP servers receive and send email addresses from all users and not only from users registered on this particular ISP server which allows the attacker sending unwanted mail. Unwanted mail usually masks its sender`s identity as well as the content of email message in order to make it more difficult to mark it as a spam. The masking can be done in these ways:

1. altering email headers – this uses badly secured user authorization and authentication implemented in SMTP protocol which creates successful mask for spam. This type creates an impression that the sender is somebody else.
2. altering email message text – this uses altering of key words in the text, for instance changing letters to numbers or, in case of email in HTML format, the text is modified in way the words are altered by inserted letters which have a font size of 0, which are present, yet not visible to the recipient. In a single campaign, one message can have several automatic modifications which makes that the same message has different length and other parameters, thus makes it difficult to antispam filters to filter these messages as a spam.
3. spam text hiding – email message body contains common informative text which is not dangerous. Text and body of the spam is included as attachment to the message as an image or another file attachment. These files contain links to web sites [2].

3.2. Unwanted email detection

As there is no clear definition of what exactly unwanted message is, it is very complicated to secure oneself against spam. That is the main reason why there is no 100 per cent protection against undesired electronic email. Having said that, sophisticated solutions are at hand to successfully filter emails.

Differentiating between unwanted electronic email and desired personal messages is basically categorization of text. For that reason, to detected spam these modified algorithms are used (specially optimized) to categorize text into two groups. The problem is in two subgroups:

3. false negative – this term is used to describe false classification of unwanted mail as wanted. This case is not determined as problematic because a small number of spam is user perfectly capable of administrating themselves.
4. false positive – this term is used to describe false classification of wanted mail as spam. As it is quite naive to expect user to control their junk mail, there is a possible chance that the user may never get to such qualified message.

To get secured against unwanted message spreading, there are several systems which use a combination of listed techniques:

- searching through email message headers to find spam characteristic features,
- analysing and validating email message sender,
- email message body scanning and finding text formatted suspiciously,
- searching for suspicious words and word formations which can be frequently found in spam messages,
- evaluating email messages by means of statistical calculations to determine probability of unwanted message.

The above mentioned techniques of protection against unwanted message spreading are implemented into systems which was a base to creation of the following methods to eliminate spam spreading:

1. Black list method – a method based on creating an IP addresses database, addresses which are misused to spread undesired mail. Providing the email message sender is listed in such a database, the message sent from this address is dropped by ISP server and thus will be delivered upon no condition.
2. Grey list method – a method based on adding unsuccessfully delivered messages to a queue. These messages can be delivered if the sender tries to send the message again. Considering fact that attackers do not try to send spam several times, this method can protect users from considerable number of junk mail.
3. White list method – this method is similar to that of black list, yet the difference is that an IP addresses database is created listing addresses that are considered safe and trustworthy.

4. Challenge-response method – This method is based on identity check when receiving a message from sender which is unfamiliar to the recipient. Email server holds the message and sends an authorization request to sender. If the sender replies, the message is delivered.
5. Email message sender verification method – this verification can be carried out in several ways:
    a. Reverse DNS Lookup – ISP server receiving the message checks if the IP address of sending ISP server is linked to domain stated in field containing sender`s address,
    b. Sender Policy Framework – writes down information about servers that do not meet requirements of Reverse DNS Lookup,
    c. Domain Keys – sending email servers ingest the outgoing mail with electronic signature which is encrypted by private key. The key by which is possible to verify the signature is stored in DNS.
6. Method of word filters and rule-based scoring systems – both methods use checking email message content to find banned phrases and key words.
7. Method of content analysis by means of algorithms – most famous algorithms for spam detection are:
    a. RIPPER algorithm – a generally defined algorithm which according to categorized samples is able to derive rules to categorize even further into deeper levels. Main advantage of this system which is able to learn is low complexity of derived rules and their altering if needed. Learning by this algorithm is separated into two parts, the limiting one and the training one. Training of this algorithm starts upon and empty set of rules. In time, the set gets larger by derived rules. An empty rule is gradually enriched with terms. Those ensure the biggest possible amount of information about objects received from test set. When the training set has no more uncovered objects, the rule derivation process ends.
    b. TF-IDF - term frequency - inverse document frequency – this algorithm is found in most algorithms for spam filtering. Differentiating between spam and desired mail is carried out by prototype vector.
    c. Bayes formula, Bayes filter – this means of filtering mail is based on mathematical and statistical model. The core of it is Bayes formula which works with probabilities. The result of creating categories of mail by Bayes filter is the probability. The algorithm determines probability of appearance of each word in set of unwanted and wanted mail. Based on gathered information, probabilities are calculated for cases when the mail is spam and contains the word or it is not a spam and includes the word nonetheless. Using Bayes formula and those probabilities, the opposite – conditioned – probability is calculated. This means the probability of cases when the message is spam and the word is present and when it is not a spam and contains the word anyway. The knowledge is stored in database as probabilities and evaluated words. To solve the problem of false positive phenomena, all probabilities are multiplied by a constant 2 and the border to filter spam is set to 0,9 [2][4].

4. UNWANTED MAIL SPREADING ELIMINATION

When spreading email marketing, it is important to lower number of email messages that are not in compliance with filters that filter messages on the side of ISP servers. As well as this, it is crucial to eliminate sending of mails to non-existing email accounts as to prevent the sender to be added on the black list containing records about such senders.

4.1. E-mail message content mail

Since the aim of a good tool to administrate email marketing is to deliver the biggest number of messages possible, it is worth checking message content before sending them using reasonable algorithms. An algorithm to check the content of email message controls it in following steps:

1. searching for forbidden words – an algorithm searches for words inside message text that filter for finding unwanted message can mark as forbidden.
2. message subject check – in this step, the algorithm checks if the subject contains capital letters which capture attention in the first place. Apart from capital letters check, algorithm controls appearance of abbreviations such as Re. or Fwd.
3. searching for colourfully formatted words – the purpose is to find colourfully formatted words in message body for colourful words tend to be part of advertising messages which aims to attract recipient`s attention. RGB, HEX values are searched for as well as colour names in HTML tags. If they affect text in any way, it is considered in the result of evaluation.
4. searching for special symbols – in this stage, the algorithm checks the text for special symbols like $, €, etc. more than once in a row, for instance €€€. In such cases, it is very probable that email message will not pass through filter.
5. links to other sites check – algorithm controls links found in the message. It checks if the letters

in link are all capital or their appearance is formatted in a very distinctive colour.

6. image to text ratio check - the algorithm tries to determine the ratio between images found in the document to actual text of the message. Images linking to social networks are ignored. If the ratio is too high, it is considered in final evaluation.
7. javascript and css code occurrence check – in this step, the occurrence of javascript code and css code is checked. Javascript code should never be found inside the message and css should not be attached in an external file but in style attribute inside HTML tags.
8. Email address and sender name check – at this stage of the processing, the algorithm checks format of the message sender. Considering email marketing, it is important for message to have sender`s address typed in with sender`s name. If the name is missing, results will be affected.
9. Email message size check – the algorithm checks the size of the message which should not exceed 30kB.
10. Cancelling subscription choice check – in the final stage of the algorithm, it is checked if one can actually cancel subscription which is decided by occurrence of key words within message body. Since subscription cancelling is an important rule of email marketing, this step has an important word to say in final evaluation [3].

This algorithm reduces possibility of spreading unwanted mail, moreover increases effectiveness of email marketing for more messages will be actually delivered to their recipients without being filtered out by spam filters.

4.2. E-mail addresses existence check

Validation of email addresses existence can be done by an algorithm that gradually sends SMTP requests to ISP server that is capable of responding to those queries.

Email address consists of a part containing identifier of an email box owner and a domain on which the box is registered.

Existence check is performed in two steps. Firstly, the existence of domain is verified and also its ability to control email addresses existence via SMTP requests. Secondly, the control of email address existence itself is performed, those are registered on these domains which are able to verify the validity of addresses.

If the connection with ISP server is set successfully, the script starts the process of validation in the following manner:

1. SMTP requests is sent, containing HELO parameter by which the script tries to establish SMTP communication with ISP server. Provided the answer is negative, the script tries to reconnect with the ISP server, yet this time it omits the HELO parameter and the parameter of EHLO is sent. This happens due to fact that some ISP servers do not support HELO parameter but do support EHLO one.
2. After the first step is completed, i.e. establishing communication, the script sends SMTP request with an MAIL FROM parameter representing sender`s email address. Since sending the request with the same sender address caused occasional troubles because the server denied to respond to loads of requests from person with the same core email address. Therefore, the address is generated automatically, thus unique each time to eliminate this problem. Such generating is allowed by gmail.com server which adding a number to valid gmail account creates a unique and valid email address.
3. In the next stage of this algorithm, SMTP request is sent, this time containing RCPT TO parameter. This parameter is created by email address which existence is being checked in this whole process or by means of which we can determine the ability of domain to reply to SMTP requests. If the server responds to this request in a positive way, this means that the email address is registered on the domain and that simply implies the existence of the address. If the opposite is true and the reply is negative, this means that the email address is not registered on this domain and is marked as nonexisting one. That is the way to find out the information about email address validity.
4. In this final step of the algorithm, after the address validation and the results obtained, it is important to restart and correctly cancel the created connection. That is done by a succession of SMTP requests, particularly RSET and QUIT that are sent over to ISP server. When the communication is over, the result is written down to a table containing results of email addresses validations.

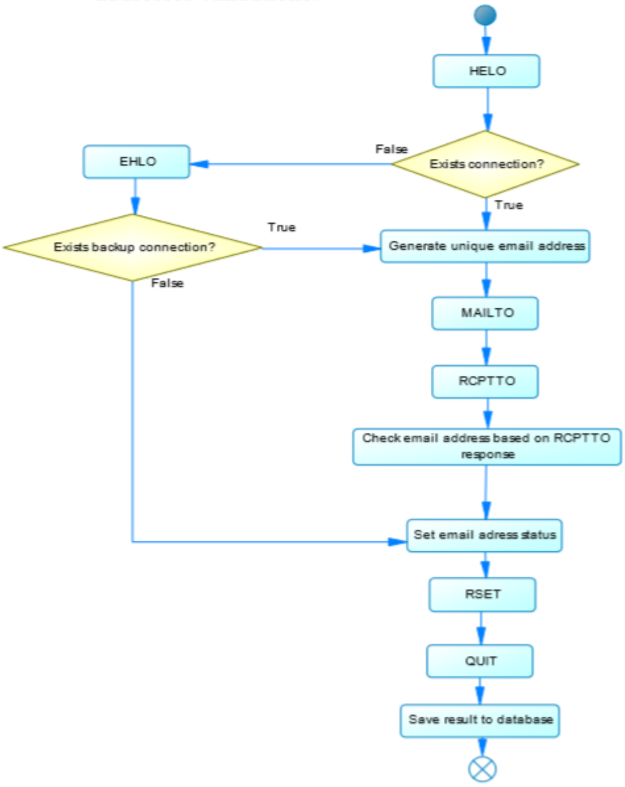

Fig. 1 Sequence of SMTP requests sending algorithm

An algorithm designed in such a way is able to eliminate sending email messages to email accounts that are invalid and therefore not registered on particular domains.

5. CONCLUSIONS

In this paper, the topics of email marketing and undesired messages spreading which is closely related to email marketing were covered in an overview fashion. In its beginning the issue of email marketing is presented as well as undesired emails. The article also contains description of a proposal on how to eliminate unwanted mail spreading by checking message content and controlling the existence of email accounts via SMTP requests to ISP servers.

ACKNOWLEDGMENTS

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0008-10 and KEGA 008TUKE-4/2013 Microlearning environment for education of information security specialists. The projects are being solved at the Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice.

REFERENCES

[1] WHITE, CH. – BAER, J.: Email Marketing Rules: A Step-by-Step Guide to the Best Practices that Power Email Marketing Success, CreateSpace Independent Publishing PlatformA247, pp 336, September 2014.

[2] ZDZIARSKI, J.: Ending Spam: Bayesian Content filtering and the Art of Statistical Language Classification., No Starch Press, 2005, pp.312.

[3] CORMACK, V.G.: Email Spam Filtering: A Systematic Review. Hanover, USA: now Publishers Inc, 2008. Pp.113. ISBN 978-1-60198-146-2.

[4] KHATER, I.M.: Hierarchical E-mail Spam Filtering Using Al & Data Mining Techniques: Decision Tree, Support Vector Machine, Multilayer Perception, Naive Bays, Bayesian Network, and Random Forests, LAP LAMBERT Academic Publishing., pp. 136, March 2012.

Received August 12, 2015 , accepted September 3, 2015

BIOGRAPHIES

Liberios Vokorokos (prof., Ing., PhD.) was born on 17. November 1966 in Greece. In 1991 he graduated (MSc.) with honours at the Department of Computers and Informatics of the Faculty of Electrical Engineering and Informatics at Technical University in Košice. He defended his PhD. in the field of programming device and systems in 2000; his thesis title was "Diagnosis of compound systems using the Data Flow applications". He was appointed professor for Computers Science and Informatics in 2005. Since 1995 he is working as an educationist at the Department of Computers and Informatics. His scientific research focuses on parallel computers of the Data Flow type. He also investigates the questions related to the diagnostics of complex systems. He is a dean of the Faculty of Electrical Engineering and Informatics at the Technical University of Košice. His other professional interests include the membership on the Advisory Committee for Informatization at the faculty and Advisory Board for the Development and Informatization at Technical University of Košice.

Ján Hurtuk (Ing.) was born on 4th October 1988 in Kežmarok, Slovakia. In 2013 he graduated (MSc.) at the Department of Computers and Informatics at the Faculty of Electrical Engineering and Informatics of the Technical University of Košice. Since 2014 he is studying as a PhD. student at the Department of Computers and Informatics at the Faculty of Electrical Engineering and Informatics of the Technical University of Košice. His scientific research is mainly focused on the computer security.

Branislav Madoš (Ing., PhD.) was born on 20th May 1976 in Trebišov, Slovakia. In 2006 he graduated (Ing.) at the Department of Computers and Informatics at the Faculty of Electrical Engineering and Informatics of the Technical University of Košice. He defended his PhD. in the field of Computers and computer systems in 2009; his thesis title was "Specialized architecture of data flow computer". Since 2010 he is working as an Assistant Professor at the Department of Computers and Informatics. His scientific research is focused on the parallel computer architectures and architectures of dataflow computers.