Network Working Group
Request for Comments: 4096
Category: Informational

C. Malamud
Memory Palace Press

May 2005

Policy-Mandated Labels Such as "Adv:" in Email Subject Headers Considered Ineffective At Best

Status of This Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Copyright Notice

Abstract

This memo discusses policies that require certain labels to be inserted in the "Subject:" header of a mail message. Such policies are difficult to specify accurately while remaining compliant with key RFCs and are likely to be ineffective at best. This memo discusses an alternate, standards-compliant approach that is significantly simpler to specify and is somewhat less likely to be ineffective.

Table of Contents

1. Labeling Requirements

The U.S. Congress and President have enacted the Controlling the Assault of Non-Solicited Pornography and Marketing Act of 2003 (CAN-SPAM Act of 2003) [US], which requires in Section 11(2) that the Federal Trade Commission:

> "[transmit to the Congress] a report, within 18 months after the date of enactment of this Act, that sets forth a plan for requiring commercial electronic mail to be identifiable from its subject line, by means of compliance with Internet Engineering Task Force Standards, the use of the characters "ADV" in the subject line, or other comparable identifier, or an explanation of any concerns the Commission has that cause the Commission to recommend against this plan."

The Korean Government has enacted the Act on Promotion of Information and Communication and Communications Network Utilization and Information Protection of 2001 [Korea]. As explained by the Korea Information Security Agency, the government body with enforcement authority under the act, Korean law makes it mandatory as of June, 2003 to:

> "include the '@' (at) symbol in the title portion (right-side) of any commercial e-mail address, in addition to the words '(Advertisement)' or '(Adult Advertisement)' as applicable. The inclusion of the '@' symbol, as proposed by the Korean government, is intended to indicate an e-mail advertisement. Because e-mails easily cross international borders, the '@' symbol may be used as a symbol for filtering advertisement mails." [KISA]

The State of Colorado has enacted the Colorado Junk Email Law, which states:

> "It shall be a violation of this article for any person that sends an unsolicited commercial electronic mail message to fail to use the exact characters "ADV:" (the capital letters "A", "D", and "V", in that order, followed immediately by a colon) as the first four characters in the subject line of an unsolicited commercial electronic mail message." [Colorado]

The Rules of Professional Conduct of the Florida Bar require, in Rule 4-7.6(c)(3) states:

> "A lawyer shall not send, or knowingly permit to be sent, on the lawyer's behalf or on behalf of the lawyer's firm or partner, an associate, or any other lawyer affiliated with the lawyer or the lawyer's firm, an unsolicited electronic mail communication

> directly or indirectly to a prospective client for the purpose of obtaining professional employment unless ... the subject line of the communication states 'legal advertisement.'" [Florida]

A subject line that complies with the above requirements might read as follows:

```
Subject: ADV: @ (Advertisement) legal advertisement
```

A more comprehensive survey of applicable laws would, no doubt, lengthen the above example considerably.

1.1. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14, [RFC2119].

2. Subject Line Encoding

The basic definition of the "Subject:" of an electronic mail message is contained in [RFC2822]. The normative requirements that apply to all headers are:

- The maximum length of the header field is 998 characters.
- Each line must be no longer than 78 characters.

A multi-line subject field is indicated by the presence of a carriage return and white space, as follows:

```
Subject: This
 is a test
```

On the subject of the three unstructured fields ("Subject:", "Comments:", and "Keywords:"), the standard indicates that these are "intended to have only human-readable content with information about the message." In addition, on the specific subject of the "Subject:" field, the standard states:

> The "Subject:" field is the most common and contains a short string identifying the topic of the message. When used in a reply, the field body MAY start with the string "Re: " (from the Latin "res", in the matter of) followed by the contents of the "Subject:" field body of the original message. If this is done, only one instance of the literal string "Re: " ought to be used since use of other strings or more than one instance can lead to undesirable consequences.

Further guidance on the structure of the "Subject:" field is contained in [RFC2047], which species the mechanisms for character set encoding in mail headers. [RFC2978] specifies a mechanism for registering different character sets with the [IANA].

In addition to choosing a character set, [RFC2047] uses two algorithms, known as "Base64 Encoding" and "Quoted Printable", which are two different methods for encoding characters that fall outside the basic 7-bit ASCII requirements that are specified in the core electronic mail standards.

Thus, an encoded piece of text consists of the following components:

- The string "=?", which signifies the beginning of encoded text.
- A valid character set indicator.
- The string "?", which is a delimiter.
- The string "b" if "Base64 Encoding" is used or the string "q" if "Quoted Printable" encoding is used.
- The string "?", which is a delimiter.
- The text, which has been properly encoded.
- The string "?=", which signifies the ending of the encoded text.

A simple example would be to use the popular [8859-1] character set, which has accents and other characters not found in the ASCII character set:

- "Subject: This is an ADV:" is an unencoded header.
- "Subject: =?iso-8859-1?b?VGhpcyBpcyBhbiBBRFY6?=" is encoded using Base64.
- "Subject: =?iso-8859-1?q?This=20is=20an=20ADV:?=" is encoded using Quoted Printable.
- "Subject: =?iso-8859-1?q?This=20is=20an=20=41=44=56=3A?=" is also encoded using Quoted Printable, but instead the last four characters are encoded with their hexadecimal representations.

Note that both character set and encoding indicators are case insensitive. Additional complexity can be introduced by appending a language specification to the character set indication, as specified in [RFC2231] and [RFC3066]. This language specification consists of

the string "*", followed by a valid language indicator. For example, "US-ASCII*EN" indicates the "US-ASCII" character set and the English language.

When a message is read, the "Subject:" field is decoded, with appropriate characters from the character set displayed to the user. Section 7 (Conformance) of [RFC2047] specifies that a conforming mail reading program must perform the following tasks:

> "The program must be able to display the unencoded text if the character set is "US-ASCII". For the ISO-8859-* character sets, the mail reading program must at least be able to display the characters which are also in the ASCII set."

However, there is no requirement for every system to have every character set. Mail readers that are unable to display a particular set of characters resort to a variety of strategies, including silently ignoring the unknown text, or generating an error or warning message.

Two characteristics of many common Message User Agents (MUAs) (e.g., mail readers) are worth noting:

- Although the subject line is, in theory, of unlimited length, many mail readers only show the reader the first few dozen characters.

- Electronic mail is often transmitted through gateways, reaching pagers or cell phones with SMS capability. Those systems typically require short subject lines.

3. Implementing a Labeling Requirement

In this section, we posit a hypothetical situation with two key players:

- John Doe [Doe] is an attorney at the firm of Dewey, Cheatem & Howe, LLC [Stooges].

- The Federal Trust Commission (FTC) has been entrusted with implementing a recent labeling requirement, promulgated by the Sovereign Government of Freedonia [Duck]. Specifically, President Firefly directed the FTC to "make sure that anybody spamming folks get the symbol ’spam:’ in the subject line and or shoot them, if you can find them."

Based on this directive, the FTC promulgated a very simple regulation which read: "Please obey the law." John Doe, being a lawyer, read the law, and promptly proceeded to spam everybody using a fairly

obvious loophole: he made sure his subject line was really long, and he shoved all the stuff like "spam:" and the "@" symbol and other verbiage near the end of the 998 allowed characters. He was complying with the law, but of course, nobody saw the labels in their reader.

Based on a periodic review, the FTC decided to be more specific, and re-promulgated their regulation as follows: "If you send spam, put 'spam:' at the _beginning_ of the subject line." The Freedonian FTC promptly received a visit from the Sylvanian Ambassador, who complained that this conflicted with his country's requirements under the Marx Doctrine to place the string "WATCH OUT! THE CONTENTS OF THIS MESSAGE ARE SUSPECT!" at the beginning of the subject line.

The re-promulgation of the regulation was rescinded, more experts were called in, and a new regulation was issued: "Put it as close to the beginning of the subject line as you can, modulo any requirements by other governments." John Doe looked at this, scratched his head, and applied a clever little hack, picking the ISO [8859-8] character set for Hebrew, and duly spelling out the letters ":" Mem Alef Pe Samech.

```
Subject: =?iso-8859-8?q?=f1=f4=e0=ee=3a?=
```

Some receivers of this message get an error message because they don't have Hebrew installed on their systems. Others get some cryptic indicator of a missing character set, such as "[?iso-8859-8?]".

The FTC called a summit of leading thinkers, and the regulation was amended to read "but don't use languages that go from right to left or up and down instead of plain old left to right." Needless to say, the reaction from the Freedonian League for the Defense of Linguistic Diversity killed that proposed regulation really quickly.

The commission continued the cycle of re-promulgation and refinement, but ultimately, the regulations continued to contain either a loophole, objectionable requirements, or violations of the relevant RFCs.

4. Subjects are For Humans, Not Labels

The use of an unknown character set, or of a very, very long subject line are just two examples of how people can try to get around labeling requirements. In order to specify a regulation without ambiguity, it would need to be extremely complex in order to avoid loopholes such as these.

Drafting of regulations is one issue, but there is another. Subject lines are used to specify, as [RFC2822] says, a "short string identifying the topic of the message."

Any regulation has to compete with the other words in the subject, and this mixing of purposes makes it very difficult for a machine to filter out messages at the direction of the user. For example, if one looks for the "@" symbol, per the Korean law, checks have to be made that this symbol is not a legitimate part of a legitimate message.

Not only do multiple labeling requirements compete with legitimate subject lines, but also there is no easy way for the sender of a legitimate message to effectively insert other labels that indicate to the recipient that-- although the message may have a required label-- it is actually a message the user might want to see, based on, for example, a prior relationship.

Even if one considers only the sender of the message, it is very difficult to specify a loophole-free way of putting a specific label in a specific place. And, even if we could control what the sender does, it is an unfortunate fact of life that other agents may also alter the subject line. For example, mailing list management software and even personal email filtering systems will often "munge" the subject line to add information such as the name of a mailing list, or the fact that a message comes from a certain group of people. Such transformations have long been generally accepted as being potentially harmful [RFC0886], and are the subject of continued discussions, which are outside the scope of the present document (see [Koch] and [RFC3834]).

The "Subject:" field is currently overloaded; it has become a handy place for a variety of agents to attempt to insert information. Because of that overloading, it is a poor location for specifying mandatory use of a label, because it is unlikely that label will "rise to the top" and become apparent to the reader of a message or even to the mail-filtering software that examines the mail before the user. The difficulty of implementing subject line labeling, without taking additional steps, has been noted by several other commentators, including [Moore-1], [Lessig], and [Levine]. Indeed, the problem is a general one. Keith Moore has pointed out seven good reasons why tags of any sort in the "Subject:" field have potential problems:

1. The "Subject:" field space is not strictly limited and long fields can be folded.

2. PDAs, phones, and other devices and software have only a limited space to display the "Subject:" field.

3. A variety of different kinds of labels such as "ADV:" and "[Mailing List Name]" compete for scarce display space.

4. There are conflicting legal requirements from different jurisdictions.

5. There is a conflict between human use of the "Subject:" field and use of that field for filtering and filing:

   * Machine-readable tokens interfere with human readability.

   * Representation of human-readable text was not designed with machine interpretation in mind and, thus, does not have a canonical form.

6. Lack of support in a few existing mail readers for displaying information from elsewhere in the message header (e.g., from newly-defined fields), along with familiarity, motivates additional uses of the "Subject:", further compounding the problem.

7. Any text-based tags added or imposed by outside parties (i.e., those that are not the sender or recipient of the message) will not be reliably meaningful in the recipient’s language.

Source: [Moore-2].

5. Solicitation Class Keywords

[RFC3865] defines the "solicitation class keyword", an arbitrary label that can be associated with an electronic mail message and transported by the ESMTP mail service, as defined in [RFC2821] and related documents. Solicitation class keywords are formatted like domain names, but reversed. For example, the registrant of "example.com" might specify a particular solicitation class keyword such as "com.example.adv" that could be inserted in a "No-Solicit:" header or in a trace field. Anybody with a domain name can specify a solicitation class keyword, and anybody sending a message can use any solicitation class keyword that has been defined by themselves or by others.

This memo argues that the "No-Solicit:" approach is either a superior alternative or a necessary complement to "Subject:" field labeling requirements because:

- One can specify very precisely what a label should be and where it should go using the "No-Solicit:" header, which is designed specifically for this purpose.

- The sender of a message can add additional solicitation class keywords to help distinguish the message. For example, if the "Freedonian Direct Marketing Council" wished to form a voluntary consortium of direct marketers who subscribe to certain practices, they could specify a keyword (e.g., "org.example.freedonia.good.spam") and educate the public to set their filters to receive these types of messages.

- Message Transfer Agents (MTAs) may insert solicitation class keywords in the "received:" trace fields, thus providing additional tools for recipients to use for filtering messages.

- A recipient can also define a solicitation class keyword, a tool that allows them to give friends and correspondents a "pass key" so the recipient's mail filtering software always passes through messages containing that keyword.

As can be seen, the solicitation class keyword approach is flexible enough to serve as a tool for government-mandated labeling and for other purposes as well.

Most modern email software gives users a variety of filtering tools. For example, the popular Eudora program allows a user to specify the name of a message header, the desired match (e.g., a wild card or regular expression, or simply a phrase to match), and an action to take (e.g., moving the message to a particular folder, sounding an alarm, or even automatically deleting messages with harmful content such as viruses). There is one popular email reader that only allows filtering on selected fields, such as "To:", "From:", or "Subject:", but that program is the exception to the rule.

In summary, for senders and receivers of email, use of the "No-Solicit:" mechanism would be simple to understand and use. For policy makers, it would be extremely simple to specify the format and placement of the solicitation class keyword. Needless to say, the issue of how to define what classes of messages are subject to such a requirement, and how to enforce it, are beyond the scope of this discussion.

6. Security Considerations

The use of labels in the "Subject:" field gives users and policy makers an unwarranted illusion that certain classes of messages will be "flagged" correctly by the MUA of the recipient. The difficulty in specifying requirements for labels in the "Subject:" field in a precise, unambiguous manner makes it difficult for the MUA to systematically identify messages that are labeled; this leads to both false positive and false negative indications.

In addition, conflicting labeling requirements by policy makers, as well as other current practices that use the "Subject:" for a variety of purposes, make that field "overloaded." In order to meet these conflicting requirements, software designers and bulk mail senders resort to a variety of tactics, some of which may violate fundamental requirements of the mail standards, such as the practice of an intermediate MTA inserting various labels into the "Subject:" field. Such practices increase the likelihood of non-compliant mail messages and, thus, threaten interoperability between implementations.

7. Recommendations

This document makes three recommendations:

1. There is widespread skepticism in the technical community that labels of any sort will be effective. Such labels should probably be avoided as ineffective at best.

2. Despite the widespread skepticism expressed in point 1, over 36 states in the U.S. and 27 countries have passed anti-spam measures, many of which require labels [Sorkin]. If such labels are to be used, despite the widespread skepticism expressed in point 1, there is a fairly broad consensus in the technical community that such labels should not be put in the "Subject:" field and should go in a designated header field.

3. If, despite points 1 and 2, a policy of mandating labels in the "Subject:" field is adopted, a complementary requirement to use the "No-Solicit:" should also be added.

8. Acknowledgements

The author would like to thank the following for their helpful suggestions and reviews of this document: Joe Abley, Harald Alvestrand, Elwyn Davies, Alain Durand, Frank Ellermann, Ted Hardie, Tony Hansen, Scott Hollenbeck, Peter Koch, Bruce Lilly, Keith Moore, Pekka Savola, and Mark Townsley.

9. References

9.1. Normative References

[IANA] IANA, "Registry of Official Names for Character Sets That May Be Used on the Internet", February 2004, <http://www.iana.org/assignments/character-sets>.

[RFC2047] Moore, K., "MIME (Multipurpose Internet Mail Extensions) Part Three: Message Header Extensions for Non-ASCII Text", RFC 2047, November 1996.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2231] Freed, N. and K. Moore, "MIME Parameter Value and Encoded Word Extensions: Character Sets, Languages, and Continuations", RFC 2231, November 1997.

[RFC2821] Klensin, J., "Simple Mail Transfer Protocol", RFC 2821, April 2001.

[RFC2822] Resnick, P., "Internet Message Format", RFC 2822, April 2001.

[RFC2978] Freed, N. and J. Postel, "IANA Charset Registration Procedures", BCP 19, RFC 2978, October 2000.

[RFC3066] Alvestrand, H., "Tags for the Identification of Languages", BCP 47, RFC 3066, January 2001.

[RFC3865] Malamud, C., "A No Soliciting Simple Mail Transfer Protocol (SMTP) Service Extension", RFC 3865, September 2004.

9.2. Informative References

[8859-1] International Organization for Standardization, "Information technology - 8-bit single byte coded graphic - character sets - Part 1: Latin alphabet No. 1, JTC1/SC2", ISO Standard 8859-1, 1987.

[8859-8] International Organization for Standardization, "Information Processing - 8-bit Single-Byte Coded Graphic Character Sets, Part 8: Latin/Hebrew alphabet", ISO Standard 8859-8, 1988.

[Colorado] Sixty-Second General Assembly of the State of Colorado, "Colorado Junk Email Law", House Bill 1309, June 2000, <http://www.spamlaws.com/state/co.html>.

[Doe] Frank Capra (Director), "Meet John Doe", IMDB Movie No. 0033891, 1941, <http://us.imdb.com/title/tt0033891/>.

[Duck] The Mark Brothers, "Duck Soup", IMDB Movie No. 0023969, 1933, <http://us.imdb.com/title/tt0023969/>.

[Florida] The Florida Bar, "Rules of Professional Conduct", 2005, <http://www.flabar.org/divexe/rrtfb.nsf/WContents?OpenView&Start=1&Count=30&Expand=4.8#4.8>.

[KISA] Korea Information Security Agency, "Korea Spam Response Center -- Legislation for Anti-Spam Regulations: Mandatory Indication of Advertisement", 2003, <http://www.spamcop.or.kr/eng/m_2.html>.

[Koch] Koch, P., "Subject: [tags] Considered Harmful", Work in Progress, November 2004.

[Korea] National Assembly of the Republic of Korea, "Act on Promotion of Information and Communication and Communications Network Utilization and Information Protection of 2001", 2001, <http://www.mic.go.kr/eng/res/res_pub_db/res_pub_mic_wp/2003/whitepaper2003/in3_7.htm>.

[Lessig] Lessig, L., "How to unspam the Internet", The Philadelphia Inquirer, May 2003, <http://www.philly.com/mld/inquirer/news/editorial/5778539.htm?1c>.

[Levine] Levine, J., "Comments In the Matter of: REPORT TO CONGRESS PURSUANT TO CAN-SPAM ACT", Federal Trade Commission, Matter No. PO44405, February 2004, <http://www.ftc.gov/reports/dneregistry/xscripts/dne040226pm.pdf>.

[Moore-1] Moore, K., "Individual Comment of Mr. Keith Moore Re: Label for E-mail Messages", Federal Trade Commission of the U.S., NPRM Comment RIN 3084-AA96, February 2004, <http://www.ftc.gov/os/comments/adultemaillabeling/040216moore.pdf>.

[Moore-2] Moore, K., "E-mail Message to the Author and the IESG", March 2005.

[RFC0886] Rose, M., "Proposed standard for message header munging", RFC 886, December 1983.

[RFC3834] Moore, K., "Recommendations for Automatic Responses to Electronic Mail", RFC 3834, August 2004.

[Sorkin] Sorkin, D., "http://www.spamlaws.com/", 2005, <http://www.spamlaws.com/>.

[Stooges] The Three Stooges, "Heavenly Daze", IMDB Movie No. 0040429, 1948, <http://us.imdb.com/title/tt0040429/>.

[US] United States Congress, "The Controlling the Assault of Non-Solicited Pornography and Marketing Act of 2003 (CAN-SPAM Act of 2003)", Public Law 108-187, 117 STAT. 2699, 15 USC 7701, December 2003, <http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=108_cong_public_laws&docid=f:publ187.108.pdf>.

Author's Address

Carl Malamud
Memory Palace Press
PO Box 300
Sixes, OR 97476
US

EMail: carl@media.org

Full Copyright Statement

Intellectual Property

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.