Internet Draft
Expiration: June 20, 2004
Network Working Group

Yakov Shafranovich
SolidMatrix Technologies, Inc.
January 20, 2004

Some of the Causes of the Spam Problem and Comparisons to Other Messaging Systems

Status of this Memo

Copyright Notice

Abstract

The problem of junk email otherwise known as "spam", has been increasing in recent years. This memo discusses some of the causes of the problem and compares it to abuse in other messaging systems.

Table of Contents

1. Introduction.

The Anti Spam Research Group (ASRG) was chartered to address the spam problem. The charter states:

> "One function of the ASRG is to look at well-specified problems that can be addressed by technical solutions."

This note falls within that category by listing some of the causes of the spam problem and comparing it to abuse problems in other messaging systems. This document is part of the work of the Inventory of Problems subgroup of the ASRG.

NOTE: This document is a product of the Anti-Spam Research Group (ASRG) of the IRTF. As per section 3 of [RFC 2014], IRTF groups do not require consensus to publish documents. Therefore readers should be aware that this document does not necessarily represent the consensus of the entire ASRG.

NOTE: This document is intended to evolve, based on comments from the Anti-Spam Research Group (ASRG) mailing list. It is certain that the current draft is incomplete and entirely possible that it is inaccurate. Hence, comments are eagerly sought, preferably in the form of suggested text changes, and preferably on the ASRG mailing list, at <asrg@ietf.org>.

2. Different Viewpoints of the Spam Problem.

Majority of today's Internet users are intimately familiar with spam. Of all forms of network abuse including viruses, DDOS attacks, and hacking, spam is the one most visible by end users. Major industry players and numerous members of the Internet community have significant dedicated time and effort to reduce and eradicate the problem, with no significant success so far.

However, a distinction is present between the viewpoints of end user and network operators. This section explore some of the aspects of the spam problem as seen from the viewpoints of different Internet players.

2.1. Aspects Relevant to End Users.

End users of the Internet look at the spam problem from a different point of view than the network operators, since they tend to see

only their own inbox. To end users spam constitutes the email they do not want to see in their inbox. Most of the time that will encompass unsolicited email - email which they did not ask for. Sometimes this will include email they did ask for but either forgot about, are too lazy to unsubscribe or granted permission under false premises.

Most end users do not care about the spam problem if it constitutes a minor part of their inbox. End users being to care about the spam problem when it grows to noticeable majority of their inbox, and begins to take away a significant amount of their time in order to deal with it.

It is important to keep in mind that ability to send an email to any email address on the Internet without prior permission is part of the open nature of the Internet. In this respect it is similar to most other human interactions where permission is not required prior to speaking or communicating with someone. Therefore, hinging the definition of spam solely upon unsolicited aspects tend to break down the open nature of the Internet, and divide it into gated communities.

2.2. Aspects Relevant to Network Operators.

The bulk aspect of spam is what also causes problems for most network operators and ISPs, to the point where most estimates at the time of writing put the amount of spam on the Internet over 50%. A small amount of spam messages as compared to the overall email traffic do not bother most network operators, since that does not significantly increase their bandwidth and personnel costs. However, when spam traffic constitutes a significant majority of email processed by network operators and ISPs, it causes major problems.

Additionally, since the spam problem involves the increased use of hijacked computers and open relays, as well as the increased number of users lacking or not willing to secure their machines properly, many network operators and ISPs begin to see increased personnel costs as they have to deal with network abuse problems that facilitate spam, rather than the spam problem itself.

These issues are compounded by the lack of cooperation among many ISPs about occurrences of network abuse, actions responding to network abuse and lack of sufficient abuse staff. This is further increased by the use of various blacklists operated by many parties, many of which are not evaluated or used properly by ISPs.

3. Some of the Causes of the Spam Problem.

Majority of the causes of the spam problem are due to the original design of the Internet itself. As described in section 3 of [IAB-E2E] the original Internet was developed as a research tool for a community of technical professionals. Thus, it did not take into account malicious behavior by its users, lack of trust between end users, and lack of authentication between end users and network operators. Additionally, the original Internet had different stakeholders which were all non-commercial, and no commercial activity was permitted on the Internet.

There are also some additional causes that have to do with the unique nature of the Internet as a communications medium. Both types of causes are described in this section.

3.1. Lack of Trust.

One major factor in the growth of spam is lack of trust between users. The Internet of the old days had few users who were non-commercial. It was also run by a single entity (the US Government), and trust among end users and end nodes was not needed. This has drastically changed with the increased number of end users and network operators, increased commercial activity, and a larger percentage of end users utilizing the Internet for malicious purposes. Due to these changes, the likelihood of a malicious email server operating on the Internet is much higher than it has been 15 years ago, which explains why the original design did not account for it.

This has lead to network operators and end users implementing different measures seeking to achieve some measure of trust. However, many of these measures reduce the ability of users to communicate without prior mutual agreement, which has been one of the major benefits of the Internet as a communications medium. An example of such measures of "challenge/response (CR)" which prevents in certain instances disabled people from communicating with others (see [TURING]).

3.2. Lack of Expertise Among End Users.

Majority of today's end users on the Internet do not possess the same technical expertise that Internet end users possessed 15 years ago. Therefore, many end users either are not interested or lack the expertise to secure their own systems. This has lead to increased occurrences of end nodes being hijacked to send spam, as well as the use of open relays to relay spam to other systems. As the result of this, network operators and ISPs have begun to take a bigger role in network security and monitoring of the network path between the end nodes, and the rest of the network.

This has lead to increased instances of well intentioned moves by ISPs that end up harming legitimate users. An example of these would be port 25 blocking which prevents legitimate roaming users from using their home MTA.

3.3. Increase in Malicious Activity.

Unlike the old Internet, a bigger number of today's users tend to utilize the Internet for malicious purposes. The original design of the Internet targeted the research community where users were unlikely to act maliciously. Additionally, since the network was smaller and more closely monitored, this has not been an issue. Due to this factor, and inherent lack of trust, some of the basic email protocols such as SMTP lack any kind of required authentication and anti-malicious activity measures.

3.4. Economic Nature of the Internet.

Communications via the Internet carries miniscule costs compared to the majority of other messaging systems available today. This is one of the reasons why the Internet has served as a disrupting technology in the economic sense, lowering costs for many

businesses. Examples of these are VoIP for telephony, online music stores for the entertainment industry, online learning for education, etc.

However, the low costs of communications, also lowered the costs of marketers, scammers and other perpetrators of illegal activities. This has allowed them to continue their activities on a much larger scale via the Internet, just like it has allowed many legitimate businesses to do the same. Being that the low cost nature of the Internet is one of the reasons why it serves so many useful purposes, it would be wrong to attack the spam problem solely from the economic angle.

3.5. Increased Role of the Third Party.

Unlike the original Internet, in today's world many third parties such as commercial entities and governments, have begun to play a larger role. This has led to a shift in focus and motivation among different users and network operators of the Internet. Since majority of ISPs and network providers are run as commercial businesses, they have an increased financial pressure to perform for their shareholders, while cutting costs as much as possible. This has led some ISPs, network operators as well as software vendors to "cut corners" when dealing with security issues and network abuse. Example of these include insufficient network abuse staff, poor security practices when writing MTA and MUA software, and increased reliance on filtering by the receiver's MTA rather than doing something on the sender's end.

The governments of the world have also begun to play a larger role on the Internet. Their interests differ greatly from the original intent of the US Government when they ran the Internet 15 years ago. Today's governments seek to provide a mutually trusted third party to facilitate trust among end users, enforce good behavior among network users, or for more malevolent reasons, enforce specific policies such as censorship.

3.6. Lack of Cooperation Among Network Operators.

Today's Internet is comprised of many more network operators than the original Internet did. Because of that, and the commercial nature of most network operators which makes them inherently compete with each other, there has been a significant reduction of cooperation among network operators and ISPs, that has been before. This has led to inability of ISPs and network operators to communicate coherently among themselves about occurrences of network abuse, and their response to it.

3.7. End to End Nature of Email.

Partly due to the end to end principle as described in [IAB-E2E], the original design of many protocols, including SMTP allows for any end nodes to the Internet to communicate and send email. This is in contrast to some of more recent IETF protocols such as [XMPP], which tend to restrict such communications to a set of authenticated nodes. However, the end to end principle has lead to the fostering of innovation on the Internet, and to a greater ability to communicate via the Internet than ever before. Therefore, we must balance this principle with the need to reduce spam (see [IAB-E2E]).

3.8. Store and Forward Nature of Email.

Another design issue that has an effect on email, is the store and forward nature of the SMTP protocol. The original Internet has many email users which were connected to it via through multiple relays, which has led to a store and forward design of SMTP, as opposed to today's point to point design of communications protocols such as [XMPP]. This makes the problem of fighting spam much harder, since it requires authentication multiple network links, as well as trusting network links to provide correct information about previous links.

However, the principle of store and forward for email, has been used for many innovative solutions on the Internet, and attacking spam solely on this principle would destroy many such solutions present today.

3.9. Social Causes.

Network abuse in general, and spam in particular, reflects the human society as whole. Just like any other business, scammers, spammers, hackers, virus writers and many others tend to utilize the Internet on a wider scale reaching a wider audience, due to the very nature of the Internet as a cheap and global communications medium. However, the very problem that causes this kind of behavior to occur is human in nature, not technological. Therefore, many social causes that play a role in the spam problem cannot be addressed by technology alone.

4. Other Messaging Systems.

While the spam problems is the most significant, many other messaging systems have similar problems as well. However, it is important to analyze why other messaging systems tend to have a lower rate of abuse problems unlike email.

This section will provide a brief comparison between email and other systems, and analyze some of the differences in regards to network abuse.

4.1. Unidirectional Systems: Television and Radio.

In unidirectional systems such as television, radio and print media, the receiver has the control over whether to participate in the message transmission or not. Receivers also have a choice of which channels or print publications to subscribe to. Thus, this ability to make a choice of which transmission to participate in, is a distinction between unidirectional systems and email. Lately, this concept has been carried over to the Internet with the recent use of RSS and a push in some newsletter circles to move over to RSS use from email. This has been done in response to the spam problem and provides receivers with an ability to choose which information they want to receive, and an ability to revoke that permission at any time. This is something that is lacking in email.

However, malicious behavior due to social issues does happen in these channels as well. Majority of such behavior involves scams and selling of products of dubious value, many of which have been transferred over to the Internet as well in the form

of spam. However, what restricts many scammers from advertising via radio, television and print media, is the discretion of the media owners, the cost of advertising and relevant laws. Additionally, advertisers in the media are usually well authenticated unlike email.

4.2. Prepaid Systems: Postal Mail and Telegraph.

In prepaid systems such as postal mail and telegraph, the problem of junk mail is present as well. However, it does not occur on such large scale as email does, due to the high cost of postage and relevant laws which forbid scams. Even though postal mail is semi-anonymous akin to email, nevertheless many legal enforcement authorities have successfully traced originators of scam mail and prosecuted them.

The same applies to telephone and fax "spam-like" abuse as well. However, telephone and faxes traveling over the telephone network, also benefit from the fact that majority of all telephone calls are authenticated end-to-end by telephone companies for billing purposes. Network operators in both mail and telephone networks, cooperate very closely both for financial and legal reasons.

4.3. Instant Messaging and SMS.

The phenomena of "spim" and spam in both IM and SMS, has been increasing lately partly due to the same social factors as email has been. However, since most IM and SMS systems are operate by a single central party, many of the problems that apply to email such as lack of cooperation between network operators, end to end principle, etc. are not applicable to IM and SMS systems. However, due to the increased convergence between IM, SMS and email, the spam problem is expected to increase within these systems. Additionally, deployment of decentralized protocols such as [XMPP] will also lead to the increase in spam on IM systems if combined with lack of cooperation between network operators, lack of security expertise among end users, and insecure deployment, implementation and configuration of XMPP and similar services.

5. Security Considerations

While much of this document deals with security issues, it does not propose any standard, and therefore does not have any direct security effects.

6. Informative References

[RFC 2026] Bradner, S., "The Internet Standards Process -- Revision 3", BCP9, RFC 2026, October 1996.

[RFC 2014] Weintrib, A., Postel, J,; "IRTF Research Group Guidelines and Procedures", BCP 8, RFC 2014, October 1996

[IAB-E2E] Kempf, J., Austein, R.; " The Rise of the Middle and the Future of End to End: Reflections on the Evolution of the Internet Architecture"; draft-iab-e2e-futures-04.txt; October 2003; (work in progress)

[TURING] May, M.; "Inaccessibility of Visually-Oriented Anti-Robot

Tests: Problems and Alternatives"; W3C Working Draft;
November 5th 2003; (work in progress)
URL: http://www.w3.org/TR/2003/WD-turingtest-20031105/

[XMPP] Saint-Andre, P,; "Extensible Messaging and Presence
Protocol (XMPP): Core"; draft-ietf-xmpp-core-21.txt;
January 6th, 2004; (work in progress)

7. Acknowledgements.

Most of the information in this note has been based on the discussions on the Anti-Spam Research Group (ASRG) mailing list. The author would like to acknowledge the contributions of all members of the group.

8. Author(s) Addresses.

Yakov Shafranovich
SolidMatrix Technologies, Inc.
research@solidmatrix.com
www.shaftek.org

9. Full Copyright Statement.