

Finding A Recipe For Spam

Chris Rose, Technology Research Institute of Florida, Inc.

Abstract

The prevalence of unsolicited e-mail, otherwise called spam, continues to haunt every user of the Internet. The overwhelming response to the government's do-not-call registry in which persons could register their telephone numbers in a database that will restrict telemarketers from calling, is an indication that people are becoming increasingly resentful of unwanted intrusions into their personal lives. It is estimated that more than a half of all e-mail, or over one trillion pieces of spam will reach the inboxes of Internet users this year but the problems of controlling spam are many since:(a) spam is virtually free for the sender (b) the SMTP protocol which governs the transmission of e-mail on the Internet was not designed to handle the complexities of deception and mistrust on a large network and (c) many major corporations are surreptitiously involved in spam. Although the development of a social conscience might keep some large corporations from engaging in spam, but spam, as we know it, would cease to exist only if either the cost of sending e-mail increased or a new secure protocol to exchange e-mail was developed. Of the two options, the quickest and easiest remedy would be to eliminate the reverse economics of sending spam by introducing a computing cost for sending e-mail.

1. Introduction

The first recorded use of the e-mail system to send unsolicited e-mail, or spam, was thought to be by a marketer for Digital Equipment Corporation (DEC), Gary Thuerk, on May 3, 1978. He was identified only as "THUERK at DEC-MARLBORO" and he decided to send a notice to everybody on the ARPANET on the west coast using a printed directory of everybody on the ARPANET as his list. The message advertised an open house at which DEC would demonstrate their new computer, the DEC-20, which was their new almost mainframe-sized system (Templeton, 2003). This was in early 1978, when the ARPANET, the forerunner to today's Internet was already providing basic e-mail services to the government departments and academic institutions that were using the network. A copy of that first spam is reproduced here:

WE INVITE YOU TO COME SEE THE 2020 AND HEAR ABOUT THE DECSYSTEM-20 FAMILY
AT THE TWO PRODUCT PRESENTATIONS WE WILL BE GIVING IN CALIFORNIA THIS
MONTH. THE LOCATIONS WILL BE:

TUESDAY, MAY 9, 1978 - 2 PM
HYATT HOUSE (NEAR THE L.A. AIRPORT)
LOS ANGELES, CA

THURSDAY, MAY 11, 1978 - 2 PM
DUNFEY'S ROYAL COACH
SAN MATEO, CA
(4 MILES SOUTH OF S.F. AIRPORT AT BAYSHORE, RT 101 AND RT 92)

A 2020 WILL BE THERE FOR YOU TO VIEW. ALSO TERMINALS ON-LINE TO OTHER
DECSYSTEM-20 SYSTEMS THROUGH THE ARPANET. IF YOU ARE UNABLE TO ATTEND,
PLEASE FEEL FREE TO CONTACT THE NEAREST DEC OFFICE
FOR MORE INFORMATION ABOUT THE EXCITING DECSYSTEM-20 FAMILY.

From: Templetons.com

Today it is estimated that 14 billion unsolicited junk e-mails or spam are sent every day and spam will cost companies \$20.5 billion in 2003. By 2007 it will cost businesses nearly 10 times that amount, or \$198 billion, to battle spam. Jupiter Research says U.S. e-mail users received more than 140 billion pieces of spam in 2001 and an estimated 261 billion pieces in 2002, which is an 86% increase. AOL says it blocks 2.3 billion spam e-mails every day and BellSouth says spam will soon add \$3 to \$5 to each customer's monthly bill (Sullivan, 2003a).

Nowadays, to be considered spam in the true sense of the word, the e-mail has to be both bulk and unsolicited. Unsolicited means that the recipient has not granted verifiable permission for the message to be sent but it can be normal e-mail such as job or sales enquiries. Bulk e-mail means that the message is sent as part of a larger group of messages, all having basically the same content but it can also be normal e-mail. These could include newsletters, discussion lists etc. but it is only when the e-mail is both unsolicited and bulk that it can be considered spam. Therefore e-mail is spam if: "(1) the recipient's personal identity and context are irrelevant because the message is equally applicable to many other potential recipients; AND (2) the recipient has not verifiably granted deliberate, explicit, and still-revocable permission for it to be sent; AND (3) the transmission and reception of the message appears to the recipient to give a disproportionate benefit to the sender" (The Spamhaus Project, 2003).

However, many consumers have a problem even deciding what e-mail really constitutes spam. For example, does a legitimate company have the right to contact a consumer and offer them a product or service? Is this spam or is it legitimate marketing? What happens if you have a previous relationship with that company are they allowed to contact you again? What is the distinction between a spammer and a direct marketing company?

2. Legislation

There have been attempts by legislators to introduce laws which would act as a deterrent to spam. At the state level, twenty-six states have anti-spam laws, and more are pending but only Delaware, bans spam outright. At the federal level, nothing has really happened although twenty spam bills, spanning everything from opt-in to wireless messaging, have been introduced in Congress. Senator Charles Schumer introduced a bill in Congress in June 2003 called the "Stop Pornography and Abusive Marketing Act" which would in effect be a "do not e-mail registry". However, only the Federal Trade Commission (FTC) has a track record of prosecuting spammers on consumer fraud charges (Lieb, 2003).

There are a number of reasons why it is much easier to introduce a do-not-call list for telephones than for e-mail. This includes the facts that the telephone industry is much more closely regulated, there are fewer corporations involved and most of them are large multinationals. These large companies have specific boundaries for their telephone connections compared to the relatively unregulated Internet in which basically anyone can rent a fast connection and then become an Internet service provider (ISP). There is also much more interconnectedness on the Internet since spam can be relayed from one server to another and more importantly, spam can come from any country in the world and be remotely controlled from any other country. Since most spam is sent by devious criminals with highly interconnected servers located all over the world, it is therefore highly unlikely that any anti-spam law would be very effective and it is doubtful that any law would deter them.

In fact, spammers sometimes hijack the computers of legitimate companies and use these computers to carry out their spamming. The computers at British Airways were hijacked by people from Argentina and used to advertise Russian mail order brides without British Airways knowing about it. This spam was thought to come from a group called Superzonda who are thought to be responsible for 20 to 30 million e-mails per day. After British Airways were alerted, Superzonda simply hijacked a server in Madrid and continued their spam business (Bomford, 2003). It is even possible for the wrong persons to be identified as the source of spam. Microsoft recently announced fifteen separate lawsuits against spammers but Simon Grainger of the United Kingdom claimed to be innocent, and claimed that he had recently purchased a domain name that might have previously been used by spammers (CNet News, 2003).

3. Reverse Economics

The economics of the Internet are also different from those of telephone networks or the postal system and this fact makes the implementation of any legislative campaign a remote possibility. In a normal network, such as the postal system, junk e-mail necessitates the sender spending time and money on the printing costs, transport, sorting, mailing etc. while at the other end the receiver can simply choose to discard the printed matter. It is true that there is a cost involved for the recipient since it takes time to dispose of the unwanted mail and there is even a cost associated with the removal of the trash. However, this cost is nowhere near the cost associated with the cost that the sender incurred while sending the mail. The Internet however, allows for the reverse economic principle of increasing returns whereby the cost of creating the first item is the largest cost and all subsequent items after the first are virtually free. For example, a software company can spend \$20 million dollars creating a piece of software, but once it has been created and hosted on their servers for download, all subsequent downloads cost the company virtually nothing. In the spam world, the initial creation of the spam organization with the corresponding computer equipment and connections are costly, but thereafter, the actual cost of sending spam is virtually nothing.

If, for example, someone has a telemarketing operation offshore, it is usually an expensive undertaking and so is the cost of direct mail which increases considerably when mail has to cross geographic borders. However, on the Internet it usually costs less to operate a sophisticated e-mail marketing program from overseas as it does from the United States. It costs a spammer basically as much to send a million e-mails as it does to send one e-mail. The economics of the business are completely weighted in favor of the sender, at the expense of the recipient. In addition, a positive response of even one-thousandth of 1% can mean good news for the sender. In other words, it takes just a few persons out of those millions of recipients to buy that body-enhancement lotion, and the advertiser has turned a profit (Microsoft b Central, 2003). It is relatively easy and incredibly cheap for anyone to send out millions of messages to anyone with an email account. Unlike any other form of marketing in history, most of the delivery costs are borne by the recipients, by the Internet-service providers (ISPs) and companies that maintain and operate mail servers and by the recipients, as anyone who has had to delete dozens of spam on a dial-up connection understands. The economics of spam are completely reversed since it is as if companies were sending us reams of postal junk mail but sending it COD (Hammond, 2003).

4. Corporations And Spam

There is a substantial grey area between spam and legitimate marketing since J.P. Morgan, Chase and Kraft U.S.A., for example, promote credit cards and coffee in ways that aren't so different from the tactics employed by anonymous peddlers of porn and gambling. "Legitimate marketers would rather the spammers disappear - but not if that means quashing the opportunity that both groups enjoy. And so the good guys let the bad guys go. It is an unspoken collusion, a sort of state-sponsored terrorism directed at our inboxes" (Hammond, 2003).

MSNBC.com recently decided to track the origins of spam messages. They clicked on a link in spam they had received and were transported to a Web page at LWSMortgage.com, where they filled out the form with traceable, fake information and waited to see what happened to their data. After four days, four companies contacted them by e-mail indicating they knew that the persons at MSNBC were looking for a new mortgage: Ameriquest, Quicken Loans, LoanWeb, and Ivy Mortgage, a small mortgage broker based in North Huntingdon, Pennsylvania. The spam did not come from any of those companies but the information came from either third party companies called "lead generators," or paid third-party contractors called "affiliates" (Sullivan, 2003b).

5. Lead Generators

There are behind the scenes Internet companies called lead generators that get lists of consumers they say are interested in, for example, new mortgages. For each package of data that they provide to a mortgage company, which would include the name, phone number, address, amount of loan desired, current home value, and other information, lead generators will earn about \$20. The mortgage company is willing to pay this \$20 since there is a potential profit of about \$1,000 profit from a new loan. Ivy Mortgage confirmed that they buy information from lead generators.

Spammers split the profits with lead generation companies thereby creating the perfect spam business since it, in effect, generates retail sales through spam. Spammers admit that it is hard to sell a product such as Viagra over the web but it is much easier to convince people to just fill out a form. The spammers also admit that they have received \$10 to \$12 per lead. Quicken Loans stated that they purchased their information from Mleads.com, a mortgage lead generation company. Further investigation by MSNBC uncovered that the initial lead came from an affiliate of Mleads, a Birmingham, Alabama company named IC Marketing. IC marketing took their data and sold it to a firm named Infoclear Marketing in Dallas, which then sold it to Mleads, which in turn sold it to Quicken Loans. This multiple layers of resale are common in the lead business and this creates a buffer between the mortgage companies and the spammers (Sullivan, 2003b).

6. Pink Contracts

The Internet Service Providers who carry the Internet spam traffic also benefit from spam. A spammer by the name of Ron Scelson filed for bankruptcy in early 2003 and the bankruptcy documents shows he owed Bell South \$56,463 for circuits and Cable & Wireless another \$4,407 as his Internet provider (Sullivan, 2003b). This is not unusual as large ISPs have previously been caught in a scheme called "pink contracts" or addendums to acceptable-use policies that allow the marketer to send spam or host a spam-related Web site on its network in exchange for "danger money" or for a higher fee because of the risk involved. A few years ago, AT&T and PSINet separately acknowledged within days of each other, the use of pink contracts that violated their respective spam policies (Olsen, 2001). In exchange for these higher rates the ISP agrees to accept more than normal complaint rates. In PSINet's contract, the firm received an upfront payment of \$27,000 from Cajunnet, a marketing firm based in Slidell, Louisiana. In exchange, PSINet agreed to permit Cajunnet to send unsolicited email "in mass quantity" through PSINet's lines (Sullivan, 2003b).

7. Technological Solutions

Many researchers have attempted to tackle the problem of spam from other angles, the most common of which is an attempt to create filters that can distinguish regular e-mail from spam. Androutsopoulos et al. (2000) did extensive research on Naive Bayesian classifiers and concluded that additional safety nets were needed for the Naive Bayesian anti-spam filter to be viable in practice. Androutsopoulos and others have also done extensive comparison work on different filtering techniques including memory-based approaches but none have returned results which would make these filters very effective against spam.

However, the work the Anti-Spam Research Group is doing is different since it has the prestige to get its proposals put in place. The group is affiliated with the Internet Engineering Task Force (IETF), which sets the standards for the fundamental technologies that make the Internet possible.

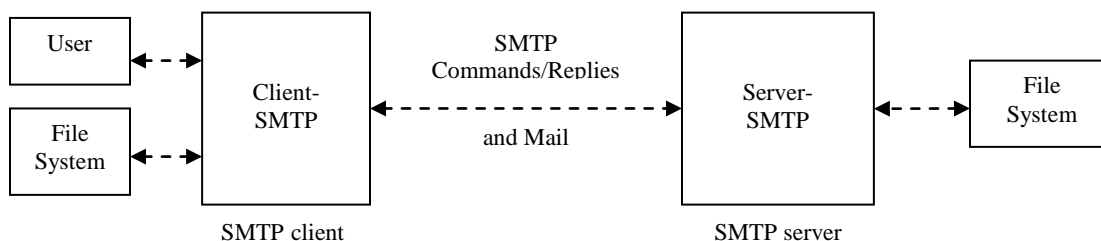
Among the technologies being standardized by the ASRG are:

- Simple authentication technology for e-mail, which will likely be implemented by Internet service providers and enterprise mail systems within several months, making it difficult for spammers to hide behind falsified sender addresses.
- "Trusted sender" technology to identify e-mail senders who can be trusted not to send spam and other unwanted e-mail.
- Reputation systems to allow everyone on the Internet to cooperate in identifying good and bad e-mail senders - similar to the reputation management systems used by buyers and sellers to rate each other on eBay.
- Interfaces for client-side tools to allow end-users to report spam, and opt out of legitimate e-mail that they no longer wish to receive.
- Developing a set of best practices for challenge-response e-mail systems, which are gaining popularity but which have the potential to create problems.
- A proposal to allow end-users to charge senders a fine using micropayments for sending unsolicited e-mail, if it turns out the end-user didn't want to get the e-mail. (Wagner, 2003)

8. The SMTP Protocol

The Simple Mail Transfer Protocol, or SMTP, assumes that you are who you say you are since it was developed when the Internet was used almost exclusively by academics. The academics and the few government departments using e-mail at that time did not have any reason to distrust one another. SMTP assumes that you are who you say you are and doesn't assume that that you're sending a Trojan horse virus, offering pills and potions to enlarge various parts of the anatomy, or that you are the relative of a deposed African dictator making a fraudulent appeal for money (Festa, 2003).

The SMTP design can be pictured as:



From: Request for Comments: 2821

If an SMTP client has a message to transmit, the protocol requires it to establish a two-way transmission channel to an available SMTP server. It is the responsibility of an SMTP client is to transfer mail messages to one or more SMTP servers, or if it cannot do that, then it should report its inability to do so.

"The objective of the Simple Mail Transfer Protocol (SMTP) is to transfer mail reliably and efficiently. SMTP is independent of the particular transmission subsystem and requires only a reliable ordered data stream channel.... An important feature of SMTP is its capability to transport mail across networks, usually referred to as "SMTP mail relaying". A network consists of the mutually-TCP-accessible hosts on the public Internet, the mutually-TCP-accessible hosts on a firewall-isolated TCP/IP Intranet, or hosts in some other LAN or WAN environment utilizing a non-TCP transport-level protocol. Using SMTP, a process can transfer mail to another process on the same network or to some other network via a relay or gateway process accessible to both networks. In this way, a mail message may pass through a number of intermediate relay or gateway hosts on its path from sender to ultimate recipient" (IETF.org, 2003).

Suzanne Sluizer co-authored the 1981 Mail Transport Protocol, SMTP's direct predecessor, while she was a technical staffer at the University of Southern California's Information Sciences Institute in Marina del Rey, California which also developed such basic Internet protocols as the Transmission Control Protocol/Internet Protocol (TCP/IP). Sluizer suggests that a new protocol to handle mail should be developed from scratch. "In my experience in computers--which at this point, is quite extensive--trying to fix problems in the existing thing is almost always more difficult than just sitting down and thinking about what you want and coming up with something new" (Festa, 2003).

Because of the simplicity and trust built into the protocol it would be necessary to introduce some level of verification into the e-mail process so as to guarantee that you are who you say you are in e-mail communications. There have been many attempts at this with varying degrees of success, including GIEIS who have developed a completely new architecture for email. It abolishes the SMTP and NNTP protocols currently in use and replaces them with a unified protocol which requires the establishment of a centralized structure for email verification. The proposed system will route all email from an ISP's customer base through a special server known as an Email Authentication Server (EAS). At this point, special encrypted codes are appended to the header that identify the account, the EAS server, and the ISP from which the email was sent. A random encrypted ID code is also appended

and these details are stored to the EAS's database. From here, the email is then distributed to the recipients EAS. The recipient's EAS then contacts the GIEIS central servers and GIEIS confirms the special encrypted codes held in the header of the email. GIEIS achieves this by communicating with the sender's EAS, confirming the database entry against codes received, deleting the database entry (or marking confirmed), and relaying the results via encrypted transmission to the recipient's EAS. This will result in one of two possible actions; either the email is sent to the recipient's email inbox or is deleted (GIEIS, 2003).

9. Economic Solutions

Other solutions have taken a different route and have focused on economic solutions to the problem of spam. Spam has the advantage that it is almost as easy to send one e-mail message as it is to send one million therefore, if a cost was added to the sending of e-mail then spammers would not be successful. This cost should not necessarily be a monetary one since all attempts at creating a true micropayment system able to efficiently handle a huge volume of micropayments have failed (Tedeschi, 2003).


There have been many attempts to introduce a computing cost to sending e-mail and Microsoft has a project called the Penny Black project which is investigating several techniques to reduce spam by making the sender pay. They have considered several currencies for payment and CPU cycles, memory cycles, and Turing tests (proof that a human was involved) are the leading candidates. Basically, this system works on the principle that if you don't know me and want to send me e-mail then you have to expend some effort to send me that e-mail. If the effort to send the e-mail is measured in CPU cycles, then since there are about 80,000 seconds in a day, a computational cost of just 10 seconds per message would limit a spamming computer to at most 8,000 messages daily. So spammers would have to invest heavily in hardware in order to send high volumes of spam (Microsoft, 2003). To achieve a ten second delay, it would be a simple matter to make a slight modification to the SMTP protocol to force each computer to compute a particular mathematical algorithm which is known to require a minimum specific amount of time to compute.

A spammer admitted to MSNBC.com that they simply had four computers and two cable modems in his operation and was able to send out 10 million e-mails a day from those computers running 24 hours a day. They had to send out about 500,000 an hour to make any money (Sullivan, 2003c) since it is estimated that the rate of return in spam is less than one-tenth of one percent (Sullivan, 2003b). A delay system such as this would force spammers to move from sending spam to the more acceptable model of sending targeted e-mail marketing messages, since instead of sending out millions of e-mail messages each day, they would be limited to thousands, unless they invested very heavily in computer equipment and Internet connections. In this particular case, instead of sending out 10 million e-mails a day this spammer would only be able to send out about 32,000, therefore it would be prudent of him to make every e-mail message count.

It is doubtful that ordinary citizens would be resentful of a system that forces a 10 second delay before they could send each e-mail if that would eliminate most, if not all of the spam that they received. Large corporations could go to a "whitelist" feature that could bypass this e-mail system. The whitelist would allow each person to set their e-mail account to allow only trusted e-mail into their system, therefore large multinational corporations that have to send out tens of thousands of e-mail messages to their employees worldwide would not be affected. However, even without a whitelist, these large corporations would be in the financial position to be able to afford the computing hardware necessary to send their corporate e-mails.

10. Conclusion

Spam has come a long way since that first message in 1978 sent by a DEC employee. In fact, spam is more prevalent than regular e-mail since more than half of all e-mail messages arriving at the inboxes of users are spam and it will cost all of us about \$20.5 billion in 2003. Legislation to prevent spam has not been very effective and because of the interconnectedness of the Internet it is doubtful if legislation can be very effective. Spam stands in stark contrast to regular junk postal mail in that it has the advantage of reverse economics whereby it costs less for the spammer to send spam than for the receiver to receive spam.

Many persons have tried technological solutions to the problem of spam mainly by the introduction of filtering mechanisms to e-mail, but these have also not been very successful. The main reason for this is that the SMTP protocol which governs the sending and receiving of e-mail was developed when the Internet was only being used by academics and the government and therefore it trusts that you are who you say you are. A technological solution would necessitate that the entire protocol be rewritten. An economic solution, on the other hand, would only require a modification to the existing protocol and an economic price in CPU cycles could be introduced. If a computing price of just 10 seconds per e-mail is implemented, then a spammer instead of sending out millions of spam per computer per day could only send about 8,000 per computer. If this computing cost was introduced then spam, as we know it, would cease to exist. 

References

1. aacug.org (2003, March 5) "Update: Antispam program stops 1 billion spams daily", Retrieved from the World Wide Web on 8/19/03 from <http://www.aacug.org/MUG/2003/aol.html>.
2. Androustopoulos, I., Koutsias, J., Chandrinou, K., Paliouras, G. and Spyropoulos, C. (2000). "An Evaluation of Naive Bayesian Anti-Spam Filtering", *Proceedings of the Workshop on Machine Learning in the New Information Age*, Barcelona, Spain.
3. Bomford, A. (2003, July 1). "Spam Peddlers Hijack Computers". *BBC.com*. Retrieved from the World Wide Web on 8/19/03 from <http://news.bbc.co.uk/2/hi/technology/3036092.stm>.
4. CNet News (2003, July 7) "Did Microsoft nab the wrong spammer?", *CNet News*. Retrieved from the World Wide Web on 8/19/03 from http://news.com.com/2009-1088_3-984352.html?tag=lh.
5. Festa, P. (2003, August 1) "End of the road for SMTP?", *CNet News*. Retrieved from the World Wide Web on 8/19/03 from http://news.com.com/2100-1038_3-5058610.html?tag=lh.
6. GIEIS website. Retrieved from the World Wide Web on 8/19/03 from <http://homepage.ntlworld.com/giza.necropolis/>.
7. Hammond, H. (2003, August). "The dirty little secret about spam", *Fast Company*. Issue 73, Page 84.
8. IETF.org (2003) "Request for Comments: 2821. Simple Mail Transfer Protocol", Retrieved from the World Wide Web on 8/28/03 from <http://www.ietf.org/rfc/rfc2821.txt?number=2821>.
9. Lieb, R. (2003, April 11). "Spam: We're Losing", *ISP Planet*. Retrieved from the World Wide Web on 8/19/03 from http://www.isp-planet.com/marketing/2003/spam_losing.html.
10. Microsoft b Central (2003). "How spam is taking money from your wallet", Retrieved from the World Wide Web on 8/20/03 from <http://www.bcentral.com/articles/tech/127.asp>.
11. Microsoft "Penny Black Project", Retrieved from the World Wide Web on 8/29/03 from <http://www.research.microsoft.com/research/sv/PennyBlack/>.
12. Olsen, S. (2001, August 19) "Giving spam the network boot", *CNet News*. Retrieved from the World Wide Web on 8/20/03 from <http://news.com.com/2100-1023-256121.html?legacy=cnet>.
13. Sullivan, B. (2003a) "Spam wars: How unwanted e-mail is burying the Internet", *MSNBC*. Retrieved from the World Wide Web on 8/27/03 from <http://www.msnbc.com/news/941040.asp?0cv=CB20>.
14. Sullivan, B. (2003b). "Who profits from spam? Surprise", *MSNBC*. Retrieved from the World Wide Web on 8/27/03 from <http://www.msnbc.com/news/940490.asp?0cb=-315171549>.
15. Sullivan, B. (2003c). "The Secret Tricks that Spammers Use", *MSNBC*. Retrieved from the World Wide Web on 9/1/03 from <http://www.msnbc.com/news/940853.asp?0cb=-415171549>.
16. Tedeschi, B. (2003, July 21). "Developing Systems of Online Payment", *New York Times*. Retrieved from the World Wide Web on 8/30/03 from <http://www.nytimes.com/2003/07/21/technology/21ECOM.html?ex=1062388800&en=41a386621c04f573&ei=5070>.
17. Templeton, B. (2003) "Reflection on the 25th Anniversary of Spam", Retrieved from the World Wide Web on 8/14/2003 from <http://www.templetons.com/brad/spam/spam25.html>.
18. "The Spamhaus Project", (2003) Retrieved from the World Wide Web on 8/14/03 from <http://www.spamhaus.org/definition.html>.
19. Wagner, M. (2003, May 25). "Major Internet Standards Group Working On Fast Plan to Can Spam", *Internet Week*. Retrieved from the World Wide Web on 8/20/03 from <http://www.internetweek.com/story/showArticle.jhtml?articleID=10100236>.

Notes