Pp. 213-218 of the *Proceedings of LISA '02: Sixteenth Systems Administration Conference* (Berkeley, CA: USENIX Association, 2002).

Over-Zealous Security Administrators Are Breaking the Internet

Richard van den Berg – Trust Factory b.v.
Phil Dibowitz – University of Southern California

ABSTRACT

As the security threats on the Internet are becoming more prevalent, firewalls and other forms of protection are becoming more commonplace. Unfortunately, improperly configured firewalls can cause a variety of problems. One particularly nasty problem is when a firewall administrator chooses to use – or continue using – Path MTU Discovery (a good choice in most situations), but blocks packets required for the protocol to work: ICMP type 3 code 4 packets. This problem, the Path MTU Discovery Black Hole, has been discussed many times before. However with under- 1500 MTU protocols such as PPPoE becoming common for both home and business high-speed connections, this problem is affecting more people than ever before.

Introduction

With the rise of security threats due to hackers, script kiddies, and viruses, the use of firewalls is becoming more widespread. This is a positive trend, but as adding firewalls to a network becomes a more common task, other problems inevitably arise. Configuring firewalls without proper knowledge of networking protocols can keep out more than one bargained for. This paper describes one common problem caused by applying overly strict packet filters incorrectly. Causes are examined and solutions are presented and analyzed.

To Filter or Not To Filter

Firewalls in their most simple form are IP routers that can be told which packets to forward and which packets to drop. This task is generally called packet filtering. Deciding what to filter and what not to filter is the hardest part of setting up a firewall. Unless the exact makeup of a network and all its IP applications is known, trial and error is the only way to find out which traffic should be allowed through. Since the idea is to increase security, denying everything unless specifically allowed is the general policy. Having a detailed knowledge of the network you are attempting to protect is critical to deploying an effective firewall.

Internet Control Message Protocol

Certain applications have proven to be more dangerous than others. In the early 90's, most vulnerabilities were found in programs like sendmail and ftpd. Filtering access to these programs was not always possible since they provided a direct service to end users. Instead, an upgrade of the software was needed to divert the attention of crackers elsewhere. In 1996 after the release of Windows 95 another type of problem surfaced. The ping of death revealed an oversight of many operating system vendors to check the validity of an Internet Control Message Protocol (ICMP) echo request packet. This caused many machines to crash. Later in 1998, the smurf attack used ICMP echo requests to flood a network by pinging a broadcast address. Since ICMP does not directly offer a service, filtering out ICMP packets seemed like a reasonable option to prevent these attacks. This completely ignored the function of ICMP in the TCP/IP suite. The main purpose of ICMP packets is error handling: letting a host know when there is a problem in the communication. The ICMP echo (ping) function can be used for debugging but is in fact far less critical.

Path MTU Discovery Black Hole

When two hosts set up a connection over the Internet using the TCP protocol, each end may let the other know what its maximum segment size (MSS) is. This MSS is derived from the maximum transfer unit (MTU) of the local interface by subtracting 40 bytes for the TCP/IP header. If somewhere along the way an IP packet does not fit in the MTU of the next link, the router handling the packet will fragment it. That is, if Path MTU Discovery is not used.

Fragmenting packets puts a strain on Internet routers, and it also degrades the overall performance of a connection. To overcome these problems, Path MTU Discovery (PMTUD) was proposed in 1988. It is now an Internet standard described in RFC 1191 [1]. PMTUD states that when two hosts communicate over TCP, the Don't Fragment (DF) bit is set. This forces a router that wants to send a large packet over a link that is too small to drop the packet and notify the sending host by sending an ICMP type 3 code 4 message. This message says the destination is unreachable, because your packet is too large and I may not fragment it. In addition to this standard ICMP message RCF 1191 adds to it: the MTU of the next link is x bytes. This way the sending host can adjust the MSS for the connection and re-send the data.

Since 1988 almost all operating systems have adopted the recommendations of RFC 1191 and use Path MTU Discovery when communicating via TCP. A problem arises when PMTUD is enabled, but incoming ICMP type 3 code 4 messages are filtered by a firewall. Since the sending host is never properly notified of any problem with the size of the packets, it will not adjust its MSS. Communication with the other host will fail. This is known as the Path MTU Discovery Black Hole problem and is described in detail in RFC 2923 [2].

The problems with PMTUD and ICMP filtering date from long before RFC 2923. One example is Path MTU Discovery and Filtering ICMP [10] explaining the issue as early as January 1998. On mailing lists like the North American Network Operators' Group (NANOG) the problem has been discussed extensively, and questions about it return every few months. This is because more and more people are being affected by the black hole.

Home and Business Networks and the Black Hole

Links with small MTU sizes are quite rare in core of the Internet. This is perhaps why filtering out all ICMP packets does not seem to cause immediate problems. However, it is causing problems with networks behind newer broadband connections in both homes and businesses (older technologies such as SLIP and X.25 are also vulnerable). Techniques like xDSL and DOCSIS (data over cable TV service) provide an Internet connection that is always on. Combined with the large bandwidth offered by these services, connecting multiple computers to the uplink becomes feasible and rewarding. The high bandwidth also requires the need of connecting over a faster medium than a serial cable (being either RS-232 or USB). Often Ethernet is chosen with PPP over Ethernet (PPPoE) as the WAN protocol. PPPoE uses encapsulation to deliver IP packets destined for the Internet to the broadband modem via Ethernet.

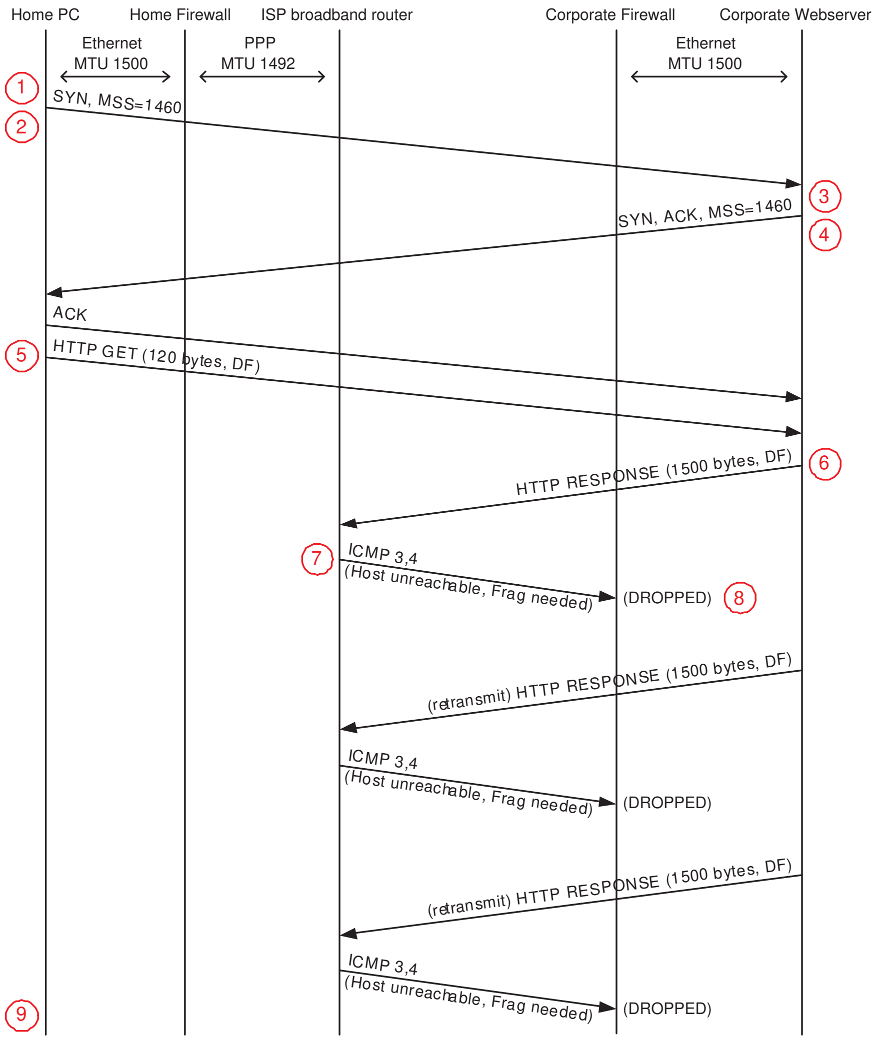

Figure 1: IP connection affected by the Path MTU Discovery Black Hole.

Going back to our Path MTU Discovery Black Hole problem, the MTU of the PPP interface will have to allow for the encapsulation so that the total PPPoE packet will fit in the standard Ethernet MTU of 1500. PPPoE interfaces therefore have a standard MTU of 1492. The disaster scenario now becomes clear:

1. A workstation on the network will start a TCP session to, say, a web server on the Internet
2. The PC sets the MSS to 1460 since the Ethernet MTU is 1500
3. The web server also connects to Ethernet, so it replies with an MSS of 1460
4. The web server enables PMTUD for the traffic to the PC
5. The PC sends an HTTP request (typically a few hundred bytes)
6. The web server starts sending the requested file, in 1500 bytes IP packets
7. The broadband router at the ISP of the end network cannot fit the packet into the PPP link and sends an ICMP type 3 code 4 message to the web server
8. A firewall between the end network and the web server drops the ICMP packet (often this is the firewall meant to protect the web server, but it can easily be any other firewall or router in between the two end networks)
9. The user is unable to browse the web site

The example uses web browsing and HTTP, but it holds true for any TCP communication sending messages of more than 1452 bytes at a time (E-mail, ftp, etc.).

Figure 1 shows a home firewall. This can either be a device specially designed for this purpose, or a generic workstation configured for this task – and could just as well be the firewall of a business. With the always-on feature of broadband connectivity, setting up a firewall at home is becoming a must.

Who Is (Not) Affected

A link with a small MTU can exist anywhere on the Internet. So theoretically everyone can be affected by this problem. As explained earlier, home and business networks utilizing PPPoE or similar protocols common in today's DSL and cable networks have a higher chance of encountering this problem. This increased probability does not apply to the following setups:

- **Just one workstation connected to a modem.** Since the MTU of the PPP interface is used to calculate the MSS for each TCP connection, the web server will only send packets that will fit into the PPP link.
- **Home gateways with a public IP address on an Ethernet interface.** The external Ethernet interface can directly be used for Internet traffic since it has its own public IP address. Since no encapsulation is needed, the MTU used for Internet traffic is 1500 (the default Ethernet MTU).
- **Home gateways connecting to a modem using USB.** Since USB does not have an MTU of its own, the PPP connection can safely use an MTU of 1500 or higher.
- **Home gateways connecting to a modem using PPTP.** The Microsoft Point-to-point Tunneling Protocol (PPTP) uses a modified version of Generic Routing Encapsulation (GRE). The MTU of the GRE interface is set to 1500. Since GRE adds 56 bytes of overhead to each packet, it is possible packets will not fit into the MTU of Ethernet. In such case, the original IP packet is fragmented even if the Don't Fragment bit is set. This is quite nasty and lowers performance [1]. It does however prevent the Path MTU Discovery Black Hole from occurring on account of the PPTP link.

Cause of the problem

As mentioned above, the Path MTU Discovery Black Hole problem is caused by using PMTUD without allowing crucial ICMP packets to pass network filters. RFC 2923 [2] describes this as an act of overzealous security administrators. It is a sign of the times to have very strict firewall policies. Check Point Software Technologies is the undisputed market leader in firewall solutions. Their FireWall-1 product used to ship with the default Policy Properties containing a setting to allow all ICMP traffic to pass. When the smurf attack hit in 1998, Check Point was publicly criticized for allowing ICMP through by default. This caused the company to change the default settings to disallow all ICMP traffic. Since Path MTU Discovery has become a standard TCP/IP feature, when anyone now installs an out-of-the box Check Point firewall, they introduce the PMTUD Black Hole. It is now left to the security administrator to explicitly allow ICMP type 3 code 4 packets to the servers that use Path MTU Discovery, or turn off PMTUD if they are uncomfortable with allowing such ICMP packets into their network.

RFC 2923 [2] mentions in Chapter 3:

> It is vitally important that those who design and deploy security systems understand the impact of strict filtering on upper-layer protocols. The safest web site in the world is worthless if most TCP implementations cannot transfer data from it.

We could not have said it any better. Many authoritative sources confirm that allowing ICMP type 3 code 4 packets through a firewall does not pose a security risk [3, 9].

Size of the Problem

If all of the above still sounds like a mere academic problem, here's the scary part: not only less experienced administrators are blocking all ICMP packets while using Path MTU Discovery. Web sites of organizations with a focus on security also have this problem. Just to name a few:

- www.securityfocus.com (recently fixed)
- www.cert.org
- www.verisign.com
- www.counterpane.com
- www.ntsecurity.com

If you cannot trust such security experts to correctly configure a firewall, whom can you trust? Is it fair to refer to this behavior as less experienced? Could not these administrators be ahead of the game by filtering something that may soon become a security hole? In fact, there is a way for administrators to successfully block ICMP type 3 code 4 packets from entering their network without breaking things. Blocking these packets without taking the proper precautions however, is not acceptable for security professionals administering firewalls. Proper solutions are discussed below.

Solutions

Since this problem has been around for quite a while, different solutions have been developed. Interestingly enough, even though the problem is caused by misconfigurations at the server side, most solutions are aimed at modifications of the clients. Apart from the moral discussion of this, it makes little sense implementation wise. If one popular server is misconfigured, all users behind a small MTU link wishing to use this server will have to adjust their settings. It would be much easier if the users could convince the maintainers of the broken site to solve the problem at its source. The truth is that this is not an easy task. This is why solving the issue at the client side is so popular.

The first three solutions we present depend on the cooperation of the (security) administrators of the websites with a misconfigured firewall. Only if this cannot be achieved, should one look at the things that can be done on the client side.

Allow ICMP Type 3 Code 4 Packets To Reach the Servers

The simplest solution is to allow Path MTU Discovery to work as it was intended: set the Don't Fragment bit on all packets and allow ICMP type 3 code 4 messages to reach the server. This means changing the overly strict rules on firewalls and other active packet filters. It should be noted that this is not considered a security risk by many authorities [3]. However, if a firewall administer feels that allowing such packets is more risk than it is worth, there are other solutions.

Disable Path MTU Discovery

If allowing ICMP into a network is not an option or cannot be achieved, the right thing to do is disable Path MTU Discovery on all servers that cannot receive ICMP type 3 code 4 packets. Since receiving these packets is a requirement for PMTUD to work [1], it breaks RFC standards and simply makes no sense to have PMTUD enabled on these servers. How to disable this feature depends on the operating system of the server. Cisco published a page with setting for some popular operating systems [4]. It is worth noting that disabling PMTUD to solve the PMTUD Black Hole will cause fragmentation. PMTUD was introduced to maximize performance by minimizing fragmentation [1]. Reintroducing fragmentation should be considered only if the previous solution is not feasible.

Path MTU Discovery Black Hole Detection

§2.1 of RFC 2923 [2] recommends the implementation of a PMTUD Black Hole detection mechanism. This is done by turning off the DF bit when retransmitting TCP packets. Various TCP/IP stacks now implement this detection scheme, but it is not turned on by default. The very nature of this solution (retransmissions) results in lower performance. Since it requires changes on the server side anyway, it makes more sense to turn off Path MTU Discovery altogether.

Using a Proxy Server

If a server is suffering from the Path MTU Discovery Black Hole, and it cannot be fixed there are some things that can be done on the client side that will prevent the Black Hole from acting up. For web browsing for example, it is possible to use a proxy server that does not suffer from the PMTUD problem. The proxy will then retrieve the pages on the client's behalf, repacking it into smaller TCP packets. Of course this only solves the problem for protocols that can be proxied.

Lowering MTU/MSS of the Internal Network

Another option is to lower the MTU of the client to the MTU of the smallest link between the client and the server. This way, the client will advertise a smaller MSS indicating to the server that its packets should not exceed this size. The same result can be achieved by lowering the maximum MSS value that a host will advertise [4]. This solution will not solve all problems. While, the MTU of the uplink is probably known and can be used as a guideline for the MTU of the systems on the LAN, one cannot be sure that this will always be the smallest MTU of the path between the clients and a server. If a smaller MTU exists on this path, ICMP type 3 code 4 messages will be sent to the server and the connection will still fail. Additionally, non-TCP protocols like UDP and IPSec will still suffer from the PMTUD Black Hole.

MSS Clamping

The solution of lowering the MTU on all systems of the LAN sounds feasible when all means less than five. If there are a dozen or more systems, this becomes a rather gruesome task. Several solutions exist to automatically adjust the MSS of TCP packets when they are being routed by the internal gateway. This is a particularly nasty solution. Per definition a router should not interfere with end-to-end settings like the MSS. Additionally, some protocols like IPSec will break when the MSS is changed in midcourse.

There are several implementations of this hack:

- --clamp-mss-to-pmtu switch for IPTables in Linux 2.4.x kernels [5]
- CLAMPMSS setting of Roaring Penguin's PPPoE Software [6, 13]
- mssfixup command of ppp for FreeBSD [7]

This solution suffers from the same problems as above: there is no guarantee that the uplink MTU is the smallest in the path (even if it is, this only works for TCP).

The MSS Initiative

In an attempt shift the focus to the cause of this issue rather than the effect, we started The MSS Initiative [8]. The purpose of this initiative is to raise awareness of systems administrators about the Path MTU Discovery Black Hole problem. We believe that when enough security administrators realize that blocking ICMP type 3 code 4 packets breaks one of the core IP protocols, they will adjust the rule sets of the devices they manage or turn off PMTUD. Gruesome hacks like MSS clamping will then become unnecessary. The MSS Initiative maintains a list of sites that are currently suffering from the Path MTU Discovery Black Hole and attempts to notify the administrators of those sites. This works in two ways: end users can check the list to see if a site they cannot reach is misconfigured, and hopefully administrators will take action upon receipt of the notice they receive from us. We also offer to help administrators unsure of how to fix their setup.

Determining if a site suffers from the Path MTU Discovery Black Hole can be difficult. It is very easy to mistake other network problems for this one. Users are encouraged to follow the instructions detailed on The MSS Initiative website if they believe a site is suffering from the PMTUD Black Hole. Users may then report the site to the Initiative so it can be added to the list and the administrator contacted.

Conclusion

Packet filters and firewalls have become necessary tools to protect systems against the growing hostility on the Internet. At the same time these tools themselves, if not configured properly, pose as a threat against one of the core protocols of the IP suite. In an ideal world, everyone would follow the guidelines set forth by Internet standards and RFCs. In a diverse and disjoint society like the Internet this cannot be expected to happen. However, when some of these standards are violated by a large number of sites and even some important vendors and security specialists fail to follow them correctly, things do break. It is in the nature of the users of the Internet to find a way around the problems that arise. Fixing things locally is attractive because of the speed and control that can be achieved, but it also allows the real problems to persist.

It is time make an effort to correct the problem that has been explained in this paper. If we do not, we might have to abandon the usage of Path MTU Discovery in the near future. This is neither efficient nor practical since it is also one of the core protocols of IPv6 [11, 12].

About the Authors

Phil Dibowitz is a junior at the University of Southern California studying Computer Engineering and Computer Science. He has held positions as a Solaris, Linux, and Netware Systems Administrator, and as a Network Administrator. His main interests are network security, networking, and UNIX. Phil maintains the official IP Filter FAQ and is the author of open source software called IPTState (monitoring software for IP Tables). Resume, projects, and accomplishments can be found at http://home.earthlink.net/~jaymzh666/ . Phil can be reached at phil@ipom.com .

Richard van den Berg has been working as a networking consultant for many large telcos and ISPs in The Netherlands since 1997. He was one of the networking specialists at Sun Microsystems' Professional Services division and currently works for Trust Factory as an independent Security Architect. Richard has the most fun decoding TCP/IP packets at the bit level while writing networking tools in various programming and scripting languages. Richard can be reached at richard@trust-factory.com .

Together the authors founded the MSS Initiative [8].

Acknowledgements

We would like to thank Chris Goggans, Marcus Ranum and the anonymous reviewers for their early reading and useful comments. We would also like to thank Joep Vesseur of Sun Microsystems for suggesting the diagram used in Figure 1. Richard likes to thank his wife Jaya Baloo for her love and support.

References

[1] Mogul, J., S. Deering, *RFC 1191 Path MTU Discovery*, http://www.ietf.org/rfc/rfc1191.txt, November 1990.

[2] Lahey, K., *RFC 2923 TCP Problems with Path MTU Discovery*, http://www.ietf.org/rfc/rfc2923.txt, September 2000.

[3] van Eden, L., *The Truth About ICMP*, http://rr.sans.org/threats/ICMP.php, May 2001.

[4] Cisco Tech Notes, *Adjusting IP MTU, TCP MSS, and PMTUD on Windows, HP and Sun Systems*, http://www.cisco.com/warp/public/105/38.shtml .

[5] Hubert, B., G. Maxwell, R. van Mook, M. van Oosterhout, P. B. Schroeder, J. Spaans, *Linux Advanced Routing & Traffic Control HOWTO*, http://www.linuxdoc.org/HOWTO/Adv-Routing-HOWTO.html, December 2001.

[6] *Roaring Penguin's PPPoE Software*, http://www.roaringpenguin.com/pppoe/ .

[7] *FreeBSD System Manager's Manual for ppp(8)*, http://www.freebsd.org/cgi/man.cgi?query=ppp&sektion=8 .

[8] Dibowitz, P., R. van den Berg, The MSS Initiative, http://home.earthlink.net/~jaymzh666/mss/, February 2002.

[9] Arkin, O., *ICMP Usage In Scanning*, Black Hat Briefings 2000, Amsterdam, http://www.sys-security.com/html/papers.html .

[10] Slemko, M., *Path MTU Discovery and Filtering ICMP*, http://www.worldgate.com/~marcs/mtu/ , January 1998.

[11] Deering, S. and R. Hinden, *RFC 2460 Internet Protocol, Version 6 (IPv6) Specification*, http://www.ietf.org/rfc/rfc2460.txt, December 1998.

[12] McCann, J. and S. Deering, *RFC 1981 Path MTU Discovery for IP version 6, August 1996*, http://www.ietf.org/rfc/rfc1981.txt.

[13] Skoll, D., "A PPPoE Implementation for Linux," *Proceedings of the Fourth Annual Linux Showcase & Conference*, Atlanta, http://www.usenix.org/publications/library/proceedings/als2000/skoll.html, October 2000.