

Path Maximum Transmission Unit (PMTU) Black Hole Routers

Joe Davies

The Internet Protocol (IP) was designed to work over an internetwork comprised of different network technologies, such as Ethernet and Frame Relay. Each network technology has a different maximum transmission unit (MTU), the maximum size of a frame that can be sent. The IP MTU is the maximum size IP packet that can be sent. A good example is Ethernet, which has an MTU of 1526 bytes. By subtracting the size of the Ethernet header and trailer (a total of 26 bytes), the IP MTU for Ethernet is 1500 bytes.

To accommodate the different IP MTU sizes of various network technologies, IP allows packets to be fragmented by routers. For example, if a packet is too large to fit on a link onto which it is being forwarded, the IP router can fragment the payload of the packet and send it as separate IP packets known as fragments.

Although this feature of IP allows for Network layer independence, it is also processor and memory intensive and can have a substantial impact on the performance of IP routers. Therefore, on modern IP networks including the Internet, fragmentation of IP packets by routers is avoided by the following:

- When sending UDP-based traffic, the maximum size of UDP messages is set low enough to prevent IP router fragmentation.
- When sending TCP-based traffic, the Don't Fragment (DF) flag in the IP header is set to 1, preventing an IP router from fragmenting the TCP segment.

When TCP peers establish a TCP connection, they exchange their TCP maximum segment size (MSS) values. The TCP peers use the smaller of the two MSS values for the TCP connection. Historically, the MSS for a host has been the MTU minus 40 bytes for the IP and TCP headers. However, support for additional TCP options, such as time stamps and selective acknowledgements, can increase the size of the typical TCP and IP header to 52 or more bytes.

When a router must fragment an IP packet and cannot because the DF flag is set to 1, it can do one of the following:

- Send an ICMP Destination Unreachable-Fragmentation Needed and DF Set message as originally defined in RFC 792 and then discard the packet. The original message format contains no information about the IP MTU of the link onto which the forwarding failed.
- Send an ICMP Destination Unreachable-Fragmentation Needed and DF Set message as redefined in RFC 1191 and then discard the packet. This new message format contains an MTU field, which indicates the IP MTU of the link onto which the forwarding failed.

RFC 1191 defines path MTU (PMTU) discovery, which allows a pair of TCP peers to dynamically discover the IP MTU, and therefore the TCP MSS, of the path between them. Upon receiving an RFC 1191-compliant Destination Unreachable-Fragmentation Needed and DF Set message, TCP adjusts its MSS for the connection to the specified IP MTU minus the TCP and IP header size so that subsequent packets sent over the TCP connection are no larger than the maximum size that can traverse the path without fragmentation.

Silently Discard the Packet

Routers that silently discard packets that needed to be fragmented but have the DF flag set to 1 are known as PMTU black hole routers.

Path Maximum Transmission Unit (PMTU) Black Hole Routers

The Cable Guy

Detecting PMTU Black Hole Routers

PMTU black hole routers can cause problems for TCP connections. For example, the TCP/IP protocol in Microsoft Windows XP and Windows Server 2003 uses PMTU discovery by default. TCP sends segments with the DF flag set to 1 and relies on the receipt of RFC 1191-compliant ICMP Destination Unreachable-Fragmentation Needed and DF Set messages containing the IP MTU to change the TCP MSS as needed.

The TCP segments exchanged during the TCP three-way handshake are not large enough to be discarded by PMTU black hole routers. However, once data begins to flow on the connection—assuming that the determined PMTU based on the negotiated MSS is larger than the actual PMTU—IP packets for TCP segments that are larger than the actual PMTU are silently discarded.

For example, you can use the FTP command line tool to successfully create a connection to an FTP server and login. However when you attempt to download or upload a file, an intermediate PMTU black hole router discards TCP segments at their maximum size, resulting in errors and an unsuccessful file transfer.

You can detect a PMTU black hole router by using the Ping tool with the following syntax:

```
ping destination -f -l ICMPEchoPayloadSize
```

- destination is an IP address or a name that can be resolved to an IP address.
- The -f option sets the DF flag to 1.
- The -l option specifies the size of the ICMP Echo message payload.
- ICMPEchoPayloadSize is the number of bytes in the ICMP Echo message payload.

To calculate the ICMPEchoPayloadSize, subtract 28 from the size of the IP packet that you want to send, because there are 20 bytes in the IP header and 8 bytes in the ICMP header of an ICMP Echo message. The following figure shows this relationship.

If your browser does not support inline frames, [click here to view on a separate page](#).

For example, to send an ICMP Echo message that is 1500 bytes long, you would use the following command:

```
ping destination -f -l 1472
```

If there are intermediate links with smaller IP MTUs and a router sends an ICMP Destination Unreachable-Fragmentation Needed and DF Set message, the Ping tool displays the message "Packet needs to be fragmented but DF set". If there are intermediate links with smaller IP MTUs and a PMTU black hole router silently discards the packet, the Ping tool displays the message "Request timed out."

To find the effective IP MTU of a path that contains PMTU black hole routers, use the Ping tool with successively larger Echo message payload sizes. Because the smallest IP MTU on typical subnets is 576 bytes, start with an ICMP Echo payload size of 548, and increase by 100 from there until you converge on the effective PMTU.

Path Maximum Transmission Unit (PMTU) Black Hole Routers

The Cable Guy

For example, if the ping 10.0.0.10 -f -l 972 command displays "Reply from 10.0.0.10" and the ping 10.0.0.10 -f -l 973 displays "Request timed out", the effective PMTU to the node assigned the IP address 10.0.0.10 is 1000 bytes (972+28).

Solutions and Workarounds for PMTU Black Hole Routers

The following solutions and workarounds for PMTU black hole routers are in order of the easiest solution to the most severe workaround.

1. Configure Intermediate Routers to Support Router-side PMTU Discovery

The easiest solution to the PMTU black hole router problem on a private intranet is to configure all your routers to support router-side RFC 1191 and the sending of ICMP Destination Unreachable-Fragmentation Needed and DF Set messages with the IP MTU of the link onto which forwarding failed. This is different than configuring a router to support host-side RFC 1191, in which the router uses PMTU discovery for its own TCP connections.

When communicating on the Internet, it is typically not possible to reconfigure Internet routers to support router-side PMTU discovery. In this case, you can use the workarounds described in the following sections.

2. Enable PMTU Black Hole Router Detection

For performance reasons, PMTU black hole router detection is disabled by default for TCP/IP in Windows 2000, Windows XP, and Windows Server 2003. If you cannot configure the routers for router-side RFC 1191 support, you can configure the following registry setting:

Setting: EnablePMTUBHDetect Key:

HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\Tcpip\Parameters

Value Type: REG_DWORD Value: 1

Because this registry entry is not present by default, you will have to add it using the Registry Editor tool and then restart Windows for the setting to take effect.

When PMTU black hole router detection is enabled, TCP tries to send segments with the DF flag set to 0 after several retransmissions of a segment are not acknowledged. If a segment with the DF flag set to 0 is acknowledged, the MSS is decreased and the DF flag is set to 1 in subsequent segments on the connection. Enabling PMTU black hole detection increases the maximum number of retransmissions that are performed for a given segment, and therefore has an effect on overall performance.

3. Determine the Best IP MTU and Set It with MTU Registry Setting

An alternative to enabling PMTU black hole detection is to determine the effective PMTU of all relevant paths by using the Ping tool as previously described in this article, and then manually configure the IP MTU of a sending interface using a registry setting.

This method avoids PMTU black hole routers by always sending IP packets with the DF flag set to 1 but at a size that does not cause the PMTU black hole routers to silently discard packets. Manually specifying the IP MTU means that all traffic will use the smaller IP MTU size, including local subnet traffic and traffic along paths that do not contain PMTU black hole routers.

After determining the effective PMTU, you can manually specify the IP MTU for a TCP/IP interface by doing the following:

Path Maximum Transmission Unit (PMTU) Black Hole Routers

The Cable Guy

- Open the Network Connections folder and note the name of the LAN connection, such as "Local Area Connection".
- Click Start, click Run, type regedit.exe, and then click OK.
- Use the tree view (the left pane) of the Registry Editor tool to open the following key: HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Control \Network\{4D36E972-E325-11CE-BFC1-08002BE10318}
- Under this key are one or more keys for the globally unique identifiers (GUIDs) corresponding to the installed LAN connections. Each of these GUID keys has a Connection subkey. Open each of the GUID\Connection keys and look for the Name setting whose value matches the name of your LAN connection from step 1.
- When you have found the GUID\Connection key that contains the Name setting that matches the name of your LAN connection, write down or otherwise note the GUID value.
- Use the tree view of the Registry Editor tool to open the following key: HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\Tcpip \Parameters\Interfaces\GUID
- Right-click the GUID key in the tree view, point to New, and then click DWORD Value.
- In the contents pane (the right pane) of the Registry Editor tool, for the value of the new registry setting, type MTU and press ENTER.
- In the contents pane, double-click the new MTU setting and in the Edit DWORD Value dialog box, click Decimal, and then type the effective MTU value in Value data.
- Click OK. Close the Registry Editor tool.
- Restart the computer for the new MTU setting to take effect.

4. Disable PMTU Discovery

If it is not possible or practical for you to determine and configure the appropriate PMTU values for each of the LAN interfaces for the Windows-based computers on your network, you can, as a last resort, disable PMTU discovery. This is not recommended because when you disable PMTU discovery, the IP MTU for all remote destinations is set to 576 bytes, which can affect performance. To disable PMTU discovery, configure the following registry setting:

Setting: EnablePMTUDiscovery Key:

**HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\Tcpip\Parameters
Value Type: REG_DWORD Value: 0**

Because this registry entry is not present by default, you will have to add it using the Registry Editor tool and then restart Windows for the setting to take effect.