计算机科学 2000Vol 27№ 8

IPsec 中 PMTU 的分析与讨论*)

The Analysis and Discussion of PMTU in IPsec

陈代兵 肖德宝
(华中师范大学计算机科学系 武汉 430079)

Abstract IPsec protocol and relative default algorithm provide high quaulity security for internet transmission. This paper introduces the notions of IPsec and PMTU. It analyzes and discusses PMTU discovery, and discusses PMTU calculation and life time and how to identify originating host in IPsec.

Keywords IPsec, PMTU, Security association, Internet security

1. 引言

中国互联网络中心发布的《中国互联网络发展状况统计报告》中有关调查表明,安全可靠性是 52.26% 的网络用户最关心的问题。此前我国网络发展速度虽然迅猛,但网络安全问题却一直没有受到足够的重视。

网络安全问题很多、很复杂,仅仅在应用层、传输层做一些认证和加密工作是远远不够的,必须在网络层下功夫。在 Internet 中,任何东西都通过 IP 传输,如何保障 IP 的安全性就是一个很大的问题。IPsec 为 IP 层提供安全服务,使系统能按需选择安全协议,决定服务所使用的算法及放置服务所需密钥到相应位置。IPsec 用来保护主机与主机间、安全网关与安全网关间、安全网关与主机间的一条或多条路径[1]。

2. IPsec 及 PMTU

目前适用于所有 Internet 通信的唯一一种安全技术就是 IPsec。IPsec 除适用于 IP 目前的版本(IPv4)外,也适用于下一代的 IP(IPv6),它提供了可交互操作的高质量的基于加密的安全。安全服务集提供包括访问控制、无连接的完整性、数据源认证、抗重播保护(序列完整性的一个组成部分)、保密性和有限传输流保密性在内的服务。这些服务是基于 IP 层的,提供对 IP 及其上层协议的保护[4]。实现这些目标,使用了两大传输协议(头部认证 AH 和封装安全负载 ESP)以及密钥管理过程与协议,而用于任何环境中的 IPsec 协议集及其使用的方式是由用户、应用程序和(或)站点、组织对安全和系统的需求来决定。

在 IPsec 结构中,由主机生成的 ICMP 消息,应该根据源 IP 地址选择符对其进行校验,这个源 IP 地址选择符与消息到达端 SA 有关。由路由器生成、受 AH 或者 ESP 保护的 ICMP 错误消息应该在 SA 隧道模式中处理和转发。没有被 AH 或者 ESP 保护的 ICMP 消息不会得到认证,该消息的处理和传送会导致此次服务的失败。一般情况下,建议不要忽略此类消息。但是可以想象许多路由器不会对传输进行 IPsec 处理,如果严格执行该规则,又会引起许多 ICMP 消息的丢弃,导致一些重要 IP 功能丢失,例如重定向和 PMTU (Path Maxinum Transfer Unit,路径最大传输单元)处理。

当一个 IP 主机向另一个主机发送大量数据时,数据作为一系列数据报传输。通常这个数据报从源到目的地的路径 的任何地方不需要分片是最好的。这个数据报大小由 PMTU 定义,并且它相当于路径中每一跳的 MTU 的最小值。IP 协议不限制数据报最小应为多小,但也不保证不分片的长数据报的可靠传递[2]。若分片了,任何一数据报片丢失,就无法重组数据报,得重传整个数据报。若将数据报定义太小(如 576 个八位组),就会有很多不必要的头部信息,降低了网络效率。因此选择一个合适的 PMTU 非常重要。当前的 IP 协议中,要发现任意路径的 PMTU 是困难的[5]。

3. PMTU 发现

下面讨论在主机和安全网关如何完成 PMTU 处理,并讨论认证的和非认证的 ICMP PMTU 消息的处理。但是,正如上面提到的,非认证的 ICMP 消息可能被本地策略丢弃。ICMP PMTU 消息携带的信息总数有限,这将影响进一步传播 PMTU 信息时哪些选择符的提供。

3.1 识别源主机

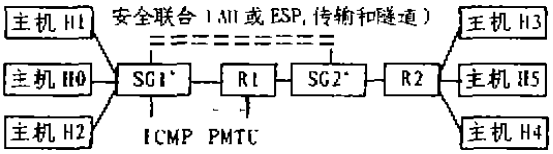
*)本项目受教育部科学技术研究重点项目资助。

ICMP 消息携带的信息量是有限的，这对可获得的选择符识别安全关联、源主机等有所影响，这些选择符用于进一步传播 PMTU 消息。简而言之，ICMP 报文必须包括来自"攻击"包的以下信息：

－IPv4(RFC 792)——IP 头加上一个 64 位最小值

相应地，在 IPv4 上下文环境里，ICMP PMTU 可能仅仅认证第一个向外的安全关联。这是因为在 IP 头之外，ICMP PMTU 仅仅容纳 64 位"攻击"报文，它只能捕获来自 AH 或 ESP 的第一个 SPI。在 IPv6 上下文环境里，ICMP PMTU 可能提供所有 SPI 和在 IP 头里的选择符，但是尽可能没有 SRC/DST 端口和封装协议(TCP、UDP 等)。而且，如果用 ESP，传输端口和协议选择符应加密。

图 1 示出了安全网关隧道的示意图(安全网关不用传输模式)。



SGx: 安全网关. Rx: 路由器. X*: X 支持 IPsec (下同)

图 1 安全网关隧道

ICMP PMTU 对 IPv4：
－Type＝3(目的地不能达到)
－Code＝4(需要分段和 DF 位设置)
－Next-Hop MTU 在 ICMP 头的第二字节的低 16 位(在 RFC 792 里，标签未用)，高 16 位置 0

IPv6(RFC 1885)：
－Type＝2(包太大)
－Code＝0(需要分段和 DF 位)
－Next-Hop MTU 在 ICMP6 的 32 位 MTU 域里

对所有在主机 H0，H1，H2 和主机 H3，H4，H5 之间的通信来说，假定 SG1 的安全策略是用单个 SA 到 SG2；并假定 H0 发送一个包到 H5，这导致 R1 发送一个 ICMP PMTU 消息到 SG1(包的内容如图 2 所示)。如果 PMTU 只有 SPI，那么 SG1 能查找 SA 并发现可能的主机名单(H0，H1，H2，通配符)；但是 SG1 没有办法确定 H0 发送触发 ICMP PMTU 消息的传输。

源包	在 IPsec 处理之后	ICMP 包
——	——	——
	IP-2 头	
	ESP 头	IP-3 头(S＝R1，D＝SG1)
IP-1 头	IP-1 头	ICMP 头(includes PMTU)
TCP 头	TCP 头	IP-2 头(S＝SG1，D＝SG2)
TCP 数据	TCP 数据	64 位 ESP hdr 最小值(＊)
	ESP 追踪者(trailer)	

图 2 数据报内容

(＊)64 位要包括足够的 ESP(AH)头，以能够容纳 SPI。

-ESP—SPI(32 位)，序列号(32 位)
-AH—下一个头(8 位)，净载长度(Payload)(8 位)，保留(16 位)，SPI(32 位)

ICMP 消息携带的信息量受限产生了一个问题，这问题在于如何为包识别源主机(以便知道进一步往哪里传 ICMP PMTU 信息)。如果 ICMP 信息仅容纳 64 位的 IPsec 头(IPv4 的最小值)，那么 IPsec 选择符(如，源和目标地址、下一协议、源和目标端口等)将丢失。但是 ICMP 错误消息仍然把 SPI、PMTU 信息和相关安全联合的源和目的地网关提供给 SG1。目的地安全网关和 SPI 只定义一个安全联合，这一安全联合反过来定义一个可能源主机的集合。按这个观点，SG1 能：

a)发送 PMTU 消息到所有可能的源主机。如果主机名单是通配符或者大多数(许多)主机不发送到 SG1，那就运行不太好。但如果 SPI/目的地等恰恰映射到一个或少量主机，就可以运行。

b)存储含有 SPI 的 PMTU，并且一直等到下一个包从相关安全联合的源主机的到来。如果它(它们)大于 PMTU，就丢掉包，用新包和更新的 PMTU 构成 ICMP PMTU 消息，把有关此问题的 ICMP PMTU 消息发给源主机。这里涉及到通知源主机的延时，但避免了 a)的问题。

既然只有后一个方法在所有场合可行，那安全网关就必须提供此类"支持"作为一个选项。但是，如果 ICMP 消息能容纳更多的来自源报文的信息，那就有足够的信息来立即确定传 ICMP 消息到哪一台主机，并且提供需要带有 5 个域(源地址、目标地址、源端口、目标端口和传输协议)的系统，以决定在哪里存储/更新 PMTU。在这种情况下，安全网关必须立即产生 ICMP PMTU 消息，这个消息在从进一步路径来的 ICMP PMTU 收据之上。注意："下一协议"(Next protocol)域可以不包含在 ICMP 消息中，ESP 加密的应用可以隐藏已经加密的选择符域。

3.2 PMTU 的计算

来自 ICMP PMTU 的 PMTU 计算必须考虑经过 H1 的任何 IPsec 头的附加—AH 和 ESP 传输，或 ESP/AH 隧道。在单个主机里，多个应用程序可以共享一个 ESP，并且安全联合的嵌套可能发生。图 3 描述了两个主机之间的安全联合的例子(与从多个主机间的一个主机的透视中看到的一样)。(ESPx 或 AHx＝传输模式)。

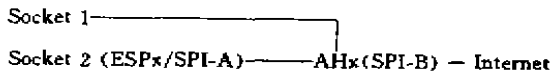


图 3 两个主机之间的安全联合

为了能够确定映射到 SPI-B 的每个套接口的

PMTU，必需有一个回调指针，指向从 SPI-B 到这两条中任一条——Socket 1 或 Socket 2/SPI-A 的路径。

3.3 维护 PMTU 数据的粒度

在主机中，做 PMTU ICMP 处理用的粒度随实现环境不同而各异。看看一台主机，关于 PMTU 问题有三种情况：

a)把 IPsec 集成到原有 IP 实现里。

b)堆栈中的块实现，在本地 IP 和本地网络设备层之间，IPsec 在已有的 TCP/IP 协议栈的实现的下面实现。

c)非 IPsec 实现—这种情况之所以存在是因为它与安全网关把信息发回主机的情况有关。

只有情况 a)时，PMTU 数据才以与通讯联合相同的粒度维护。其它情况下，IP 层在源和目的 IP 地址的粒度里维持 PMTU 数据。这点是重要的区别，因为不止一个通讯联合可以映射到源和目的 IP 地址，在每个联合之上可以有不同数量的 IPsec 头(例如，由于用不同的转换或不同的算法)。下面的例子阐述了这种情况。

在情况 a)和 b)里，假设你有如下条件：H1 正发数据到 H2，并且从 R1 传送到 R2 的包超过它们之间的网络跳的 PMTU。图 4 示出了嵌有 IPsec 的主机间的数据传输。

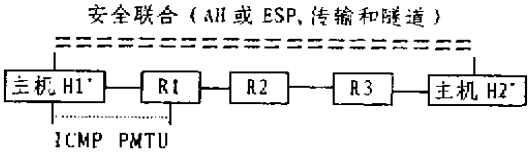


图 4 嵌有 IPsec 的主机间的数据传输

如果 R1 配置成不分段的签署者传输(subsciber traffic)，那么 R1 发送带有适当的 PMTU 的 ICMP PMTU 消息给 H1。H1 的处理与实现有本质不同。在情况 a)下(原有 IP)，安全服务绑定在套接口或其对等实体上。这里 H1 里的 IP/IPsec 实现能够为联合套接口存储/更新 PMTU。在情况 b)下，H1 的 IP 层能够存储/更新 PMTU，但是如上面谈到的一样，只在源和目的地址的粒度和可能的 TOS/Class 上。因此这结果是次最佳的，因为给定了源/目的/TOS/Class 的 PMTU 是最大数量的 IPsec 头的减少，这个头用于给定的源和目的地之间的所有通信联合。

在情况 c)下，必须有一个安全网关来做任何的 IPsec 处理。因此假定你有下面的条件：H1 正发数据到 H2，并且从 SG1 到 R 的包超过它们之间的网络跳的 PMTU。图 5 示出了用安全网关来做 IPsec 处理的数据传输。

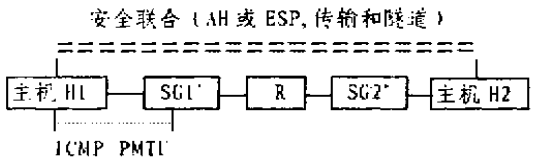


图 5 用安全网关来做 IPsec 处理的数据传输

如上面描述的情况 b)一样，H1 的 IP 层能够存储/更新 PMTU，但只在源和目的地址的粒度和可能的 TOS/Class 上。因此这结果是次最佳的，因为给定了源/目的/TOS/Class 的 PMTU 是最大数量的 IPsec 头的减少，这个头用于给定的源和目的地之间的所有通信联合。

3.4 PMTU 的每个套接口维护

PMTU 计算的实现和以单个"通信联合"的粒度对 PMTU 的支持都是本地事件。但是在主机中的基于套接口的 IPsec 实现应该把信息维持在每个套接口基础上。堆栈中的块系统必须在 ICMP PMTU 之后，把它传送给主机 IP 实现，这个 PMUT 用来调整这些系统增加的 IPsec 头。其上的决定应该由返回的 ICMP PMTU 里的 SPI 分析和当前别的选择符信息来确定。

3.5 PMTU 数据的生命期

在所有实现 IPsec 和保留 PMTU 信息的系统中，与安全联合有关的 PMTU 必须"计时"，而且为了及时更新 PMTU 也需要采取一些机制，特别是如果想发现 PMTU 但 PMTU 比所需的要小的时候。一个给定的 PMTU 必须保留足够长的时间，以将从安全联合源端得到的包传给安全联合另一端的系统，如果当前 PMTU 太大则传回一个 ICMP 错误消息。注意如果有嵌套隧道，使 ICMP 消息传回一个封装者或者源主机可能需要多个包和循环传输次数。系统应该使用 Path MTU 发现路径，它建议定期将 PMTU 重设为第一跳的数据链接的 MTU，并根据需要让正常 PMTU 发现过程更新 PMTU。

参 考 文 献

1 Kent，Atkinson. Security Architecture for IP. RFC2401，1998

2 Mogul，Deering. Path MTU Discovery. RFC1191，1990

3 Alexei V. Top View on the IPsec Technology. 1998

4 Doraswamy N，Harkins D. IPsec：the new security standard for the Internet，intranet，and virtual private networks. 1999

5 Comer D E. 著，林瑶等译. TCP/IP 网际互连. 1998

6 Atkins D. Internet 网络安全专业参考手册