组织：中国互动出版网（ http://www.china-pub.com/ ）

RFC 文档中文翻译计划 (http://www.china-pub.com/compters/emook/aboutemook.htm)

E-mail：ouyang@china-pub.com

译者：徐国栋（ xgdong_y994 xgdong@sina.com ）

译文发布时间： 2001-6-27

Network Working Group
Request for Comments: 2475
Category: Informational

S. Blake
Torrent Networking Technologies
D. Black
EMC Corporation
M. Carlson
Sun Microsystems
E. Davies
Nortel UK
Z. Wang
Bell Labs Lucent Technologies
W. Weiss
Lucent Technologies

December 1998

分类业务的体系结构

(An Architecture for Differentiated Services)

本文档的状态

本文档为互连网社区提供一般性的知识。并未定义任何互连网标准。对本文档资料的分发、传播不受限制。

版权声明

摘要

本文档定义了一种可以在互连网上实现可扩展的分类业务的体系结构。这种体系结构通过标记 IP 层数据包的 DS 段[DSFIELD]，体现不同的业务级别，从而提供扩展性业务。在一个数据包的传输路径上的每一节点，都根据该数据包的分类标记为其提供特定的传输服务。复杂的分类，标记，传输策略，及整形操作仅仅需要在网络边缘或用户主机上实现。网络资源根据服务策略而被分配给不同的业务流。这些服务策略管理着业务数据在进入一个具有分类业务能力的网络时，如何标记，调整，并在网络中向前传输。在这些基本分类业务模块的基础上，可以实现各种各样的服务。

目 录

1 介绍

1.1 综述

本文档定义了一种可以在互联网上提供可扩展的分类业务的体系结构。一种“业务”，是由在一个网络内，在同一个传输方向上，通过一条或几条路径传输数据包时的某些重要特征所定义的。这些特征可能由吞吐率，时延，时延抖动，和 /或丢包率的量化值或统计值所指定，也可能由其获取网络资源的相对优先权来指定。业务分类要求能适应不同应用程序和用户的需求，并且允许对互联网服务的分类收费。

本体系结构由许多在网络节点上实现的功能实体组成，包括每一跳转发行为集合，数据包分类功能，和业务量调节功能。其中，业务量调节功能又有测量，标记，整形，和监察四部分。在本体系结构，只在网络的边界节点上实现复杂的分类和调节功能。并且，通过在 IPv4 和 IPv6 包头的 DS 段做适当的标记 [DSFIELD] ，把业务量归为集合，然后根据所做的标记，采取不同的每一跳转发策略。因此，本体系结构具备可扩展性。“每一跳行为”保证了在每个网络节点，为互相竞争资源的业务流分配缓冲区和带宽资源时，有一个合理的处理粒度。在核心网络节点上，为每个应用程序业务流或者为每个用户维护一个转发状态是不可行的。在以下功能中是有区别的：

- 向业务集合提供的服务
- 用于实现某种服务的调节功能和每一跳行为
- 用于标记数据包从而选择每一跳行为的 DS 段值（ DS 编码点）
- 实现每一跳行为时，特定节点的实现机制

在网络内部节点，服务提供和业务量调节策略被有效地同数据包转发策略分离开。这样，保证了网络可以提供相当广泛的服务类型，并给未来的扩展留下足够的空间。

本体系结构只在一个业务流方向上提供分类业务，它是非对称的。开发出一种对称式的体系结构是目前研究的一个课题，但已经超出了本文档的描述范围；感兴趣的读者可以参考 [EXPLICIT] 。

1.2 节是本文档使用的术语表。 1.3 节列出了本体系结构所解决的需求。 1.4 节提供了与其它分类业务解决方案的简要比较。第 2 节详细介绍了本体系结构中的各个模块。第 3 节建议了每一跳行为规范的设计准则。第 4 节讨论了与没有实现本文档及 [DSFIELD] 定义的分类业务功能的节点和网络的互操作问题。第 5 节讨论了与多点传送有关的问题。第 6 节讨论安全和隧道问题。

1.2 术语

本节给出了在本文档中所使用术语的一般性概念解释。其中的某些术语将在文档后面章节中给出更准确的解释。

行为集合（ Behavior Aggregate : BA ）	一个 DS 行为集合。
BA 分类器（ BA Classifier ）	仅基于 DS 段的内容选择数据包的分类器。
边界连接（ Boundary Link ）	连接两个域的边界节点的连接。
分类器（ Classifier ）	根据已定义的规则和包头内容选择数据包的实体。

DS 行为集合（ DS Behavior Aggregate ）	在一个特定方向上，通过一条连路传输的具有相同 DS 编码点的数据包集合。
DS 边界节点（ DS Boundary Node ）	在 DS 域中，负责连接另一个ＤＳ域或者连接一个没有ＤＳ功能的域的节点。
具有ＤＳ功能（ DS-capable ）	实现了本体系结构中定义的分类业务功能的；通常用于形容一个由ＤＳ兼容节点组成的域。
ＤＳ编码点（ DS Codepoint ）	ＤＳ段中ＤＳＣＰ部分的特定值，用于选择ＰＨＢ。
ＤＳ兼容（ DS-compliant ）	能够支持在 [DSFIELD] ，本文档，和其它有关分类业务的文档中定义的分类业务功能的；通常用来形容一个节点或者网络设备。
ＤＳ域（ DS Domain ）	具有ＤＳ功能的域；连续分布的节点的集合，它们具有共同的服务提供策略和ＰＨＢ定义。
ＤＳ出口节点（ DS Egress Node ）	处理离开此ＤＳ域的业务流的ＤＳ边界节点。
ＤＳ入口节点（ DS Ingress Node ）	处理进入此ＤＳ域的业务流的ＤＳ边界节点。
ＤＳ内部节点（ DS Interior Node ）	非边界节点的ＤＳ节点。
ＤＳ段（ DS Field ）	在ＩＰｖ４中，指ＴＯＳ字节；在ＩＰｖ６中，指业务类型字节。其中的ＤＳＣＰ段诸比特用于编码ＤＳ编码点，其它的比特目前没有使用。
ＤＳ节点（ DS Node ）	ＤＳ兼容的节点
ＤＳ区（ DS Region ）	连续分布的ＤＳ域的集合，在其上可以建立跨越多个ＤＳ域提供分类业务的连路。
下游ＤＳ域（ Downstream DS Domain ）	一个边界连接中，位于业务流下游的ＤＳ域。
丢包器（ Dropper ）	负责丢包的功能模块。
丢包（ Dropping ）	基于一定的原则丢弃数据包；参见 监察（Policing ）。
遗留节点（ Legacy Node ）	实现了在 [RFC791 ， RFC1812] 中定义的 IPv4 优先算法，但并非 DS 兼容的节点。
标记器（ Marker ）	负责标记的功能模块。
标记（ Marking ）	基于一定的原则设置一个数据包的 DS 编码点；参见 预标记（ Pre-marking ），重标记（Re-marking ）。
机制（ Mechanism ）	在节点中用于实现一种或多种每一跳行为的特殊算法或操作（例如，排队策略）。
测量器（ Meter ）	负责测量的功能模块。

测量（ Metering ）	计算由分类器选中的业务流的时间性特征（例如，速率）。这一过程的即时状态可能会影响标记器，整形器，或者丢包器的行为，也可能被用于记帐收费或者纯粹的测量目的。
微流（ Microflow ）	一个独立的从应用程序到应用程序的数据包流，由源地址，源端口号，目的地址，目的端口号和协议标识符区分。
MF 分类器（ MF Classifier ）	根据任意数目的包头字段的内容来选择数据包的多字段（ MF）分类器。典型的字段组合可能包括源地址，目的地址， DS 段，协议标识符，源端口号和目的端口号。
每一跳行为（ Per-Hop-Behavior : PHB ）	在 DS 兼容节点上，作用在 DS 行为集合上的外界可观察的转发行为。
PHB 组（PHB Group ）	由一个或多个 PHB 组成的集合。这些 PHB 由于共同的限制，例如队列服务或队列管理策略，必须同时被指定及实现。 PHB 组提供了构建服务的基石，使得一系列的转发行为可以被同时指定。一个单独的 PHB 是 PHB 组的特例。
监察（ Policing ）	根据依照某种业务量简档工作的测量器的状态，丢弃（通过丢包器）业务流的部分数据包。
预标记（ Pre-mark ）	在数据包进入下游 DS 域之前，设置其 DS 编码点。
提供者 DS 域（Provider DS Domain ）	具有 DS 功能的服务提供者所属的源域。
重标记（ Re-mark ）	改变数据包的 DS 编码点。通常由标记器根据 TCA 确定如何修改。
服务（ Service ）	在 DS 域内或者在端到端条件下，对用户业务量的一个确定的子集所采取的所有处理措施。
服务水平协议（ Service Level Agreement ： SLA）	用户和服务提供者之间达成的关于如何为用户提供转发服务的服务协议。这里的用户可能是一个使用者组织（源域），也可能是另一个ＤＳ域（上游域）。服务水平协议 SLA 可以包括部分或全部组成一个ＴＣＡ的业务量调节规则。
服务提供策略（ Service Provisioning Policy ）	关于业务调节器如何配置到ＤＳ边界节点上，及业务流如何映射到特定的ＤＳ行为集合以获得某些服务的策略。
整形器（ Shaper ）	负责业务量整形的功能模块。
整形（ Shaping ）	有意延迟业务流中的某些数据包，以使业务流符合预先定义的业务量简档。

源域（ Source Domain ）	发出接受某种特定服务的业务流的节点所在的域。
业务量调节器（ Traffic Conditioner ）	负责完成业务量调节功能的功能实体。包括测量器，标记器，丢包器，和整形器。业务量调节器可以重新标记业务流，或者丢弃或整形数据包，从而改变业务流的时间特征，使业务流符合事先达成的业务量简档。
业务量调节（ Traffic Conditioning ）	实现ＴＣＡ中确定的控制规则，包括测量，标记，整形，和监察。
业务量调节协议（ Traffic Conditioning Agreement：TCA ）	一份指明应用到分类器选中的业务流的分类规则，相应的业务量简档，以及对此业务流的测量，标记，丢弃，和/或整形规则的协议。ＴＣＡ包括来自三方面的业务量调节规则：ＳＬＡ显式指定，相关的服务需求隐式指定，和/或来自于ＤＳ域的服务提供策略。
业务量简档（ Traffic Profile ）	关于业务流的时间特征的描述，例如速率和突发包大小。
业务流（ Traffic Stream ）	具有管理重要性的通过同一段路径的一个或多个微流的集合。业务流可能包含由特定的分类器选出的活动的微流集合。
上游ＤＳ域（ Upstream DS Domain ）	一个边界连接中，位于业务流上游的ＤＳ域。

1.3 需求

在互联网的发展历史上，从主机数目，到应用程序的种类和数量，再到网络基础设施的能力，都有着持续的增长。而且，这种增长在可预见的未来还会持续。因此，必须有一种支持分类业务的可扩展体系结构与这种持续增长相适应。

在这种体系结构中，下列需求必须得到认可，并能被满足：

- 提供从端到端或者在特定网络（或网络集合）内部的，多种多样的服务和提供策略。
- 允许将服务从特定的应用程序中分离出来。
- 能够与已有的应用程序共存，而无须改变应用程序编程接口或者主机软件（假设适当配置了分类器，标记器，和其它的业务量调节功能模块）。
- 应该在核心网络节点实现时，将业务量调节和服务提供功能同转发行为相分离。
- 不应依赖逐跳的应用程序信令。
- 仅需要一个很小的转发行为集合。其实现复杂性不应是网络设备开销的主要部分，也不应给未来高速系统的实现引入瓶颈。
- 应该避免在核心网络节点内为每个微流或者每个用户保持各自的状态。
- 在核心网络节点内，应仅保存集合分类状态。
- 允许在核心网络节点实现简单的数据包分类（ BA 分类器）。

- 允许同无 DS 兼容性的网络节点的合理的互操作性。
- 具备增量式布署能力。

1.4 和其它方法的比较

在本文档中定义的分类业务体系结构可以同其它已存在的分类业务模型相比较。我们把这些可选的模型分为如下几类：相对优先级标记，服务标记，标签交换，集成业务 /RSVP，和静态逐跳分类。

相对优先级标记模型的例子包括 [RFC791] 定义的 IPv4 优先级标记， 802.5 令牌环优先级 [TR]，和缺省的 802.1p 业务量分类 [802.1p] 。在这个模型中，应用程序，主机，或者代理节点为数据包选择一个相对优先级（例如，延迟或者丢弃优先级）。在整个传输路径上的网络节点根据包头中指定的优先级采取相应的转发行为。我们的体系结构可以被认为是这种模型的更新。在这种体系结构中，更清楚的指明了边界节点和业务量调节器的作用及重要性；并且，每一跳行为模型也允许比相对延迟或丢弃优先级更具一般性的转发行为。

服务标记模型的一个例子是 [RFC1349] 定义的 IPv4 TOS 。在这个例子中，每个数据包被标记为需求某种“服务类型”，包括“延迟最小化”，“吞吐量最大化”，“可靠性最大化”，或者“费用最小化“。网络节点根据标记的服务类型选择路由或者转发行为。这个模型同我们的体系结构有细微的差别。请注意，我们并没有描述使用 DS 段做为路由选择的输入。[RFC1349] 定义的 TOS 标记具有广泛的一般性，无法扩展可能的服务语义范围。而且，其服务需求是与每一个数据包相关联的，但有些服务语义可能依赖于一系列数据包的整体转发行为。服务标记模型不能很容易的适应未来服务范围和数量的增长（鉴于其编码空间太小），而且在每一个核心网络节点都会涉及“ TOS 到转发行为”的转换。服务标记的标准化还意味着提供服务的标准化，这已经超出了 IETF 的工作范围。注意服务提供记录在分配的 DS 编码空间中，从而允许具有本地重要性的编码点被提供者用于提供服务标记语义 [DSFIELD] 。

标签交换（或叫做虚电路）模型的例子包括帧中继， ATM，和 MPLS[FRELAY ，ATM] 。在这种模型中，沿网络路径的每一跳，都建立业务流的路径转发状态和业务管理或 QoS 状态。各种不同粒度的业务量集合在入口节点处与一条标签交换路径相关联。在每一标签交换路径内，数据包或信元被赋予一个转发标签。转发标签负责寻找下一跳节点，每一跳转发行为，和在每一跳时的标签置换。由于标签并非全局性的，而只是在一条链路上有效，所以这种模型允许对业务量分配资源时能采取更好的粒度。也正因为如此，网络资源可以被预留给在某条链路上收到的具有特定标签的数据包或信元集合，同时，标签交换语义控制着下一跳路由选择，允许业务流通过特别设计的路径穿过网络。这种改进的粒度控制是以增加建立和维护标签交换路径的管理和配置需求为代价的。并且，在最好情况下，每个节点保存的转发状态数量与边界节点数量成正比（假设存在多点到点的标签交换路径）；在最坏情况下（采用提供资源的边到边标签交换路径），与边界节点数量的平方成正比。

集成业务 /RSVP 模型在缺省情况下依赖传统方式转发数据包，同时，它也允许发送方和接收方通过信令交互在两者之间的路径上每个节点处建立额外的数据包分类和转发状态 [RFC1633 ，RSVP] 。由于缺少对业务流的归类，每个节点保存的状态数将与并发的资源预留数成正比。在一些高速链路上，这个数目可能会很大。这个模型还需要应用程序支持资源预留信令协议。在核心网络节点，可以使用分类业务机制将集成业务 /RSVP 状态归类 [BERNET] 。

集成业务 /RSVP 模型的一个变种通过在网络路径沿途的每个节点处只采用“静态”分类和转发策略，使逐跳进行信令交互变的不再需要。这些策略是管理级的，并非针对网络中的活动微流。这个变种的状态需求可能会比 RSVP 更多，特别是在骨干网节点处。因为随着时间推移，一个节点所采用的静态策略数可能比在此节点请求资源预留的活动的发送 -接收对话数还要

多。虽然采用大数量的分类规则和转发策略在计算复杂性上可行，但由此而需要在业务流必经的骨干网节点处安装和维护这些规则的管理负担也是需要认真考虑的。

以上把我们提出的体系结构与其它的分类业务模型进行了比较。需要注意的是，采用这些技术的链路和节点应该是通过基于第二层交换的网络结构（例如，802.1p 局域网，帧中继/ATM 骨干网）互连 DS 节点，来提供分类业务行为和语义。对于 MPLS（多协议标签交换）条件下，可以作为可选的域内实现技术。在 DS 域（或者在提供 DS 域接入的网络内）的特定区域采用特殊的链路层技术，意味着对业务流更粗粒度的分类。依赖于从 PHB 到不同的链路层服务的映射和把数据包安排到有限优先级（或者不同类型和能力的虚电路）的方式，全部或部分使用中的 PHB 是可被支持的（或者是不可辨别的）。

2 分类业务体系结构模型

分类业务体系结构基于这样一个简单模型：进入网络的业务量在网络边缘处进行分类和可能的调节，然后被分配到不同的行为集合中去。每一个行为集合由唯一的 DS 编码点标识。在网络核心处，数据包根据 DS 编码点对应的每一跳行为转发。在本节中，我们讨论在分类业务区域中的关键组件，业务量分类和调节功能，以及分类业务是如何通过业务量调节和基于 PHB 的转发而实现的。

2.1 分类业务域（DS 域）

DS 域是邻接的 DS 节点集合。这些 DS 节点执行共同的服务提供策略，并实现相同的 PHB 组。每个 DS 域都拥有完好定义的边界。位于边界处的 DS 边界节点负责将进入此 DS 域的业务流分类及进行可能的调节，以保证穿过此 DS 域的数据包被适当标记，并按照 DS 域所支持的 PHB 组中的一个 PHB 转发。DS 域内的节点根据 DS 编码点为数据包选择转发行为。从 DS 编码点值到某个被支持的 PHB 组的映射，依赖的是推荐的编码点到 PHB 的映射规则或者用户定义的本地化映射规则 [DSFIELD] 。如果在 DS 域中包含非 DS 兼容节点，那么很可能导致性能表现的无法预测，并且会妨碍服务水平协议（SLA）的实现。

一个 DS 域通常包含一个或多个处于同一组织管理下的网络；例如，一个组织的内部网或者一个 Internet 服务提供商（ISP）。域管理者必须保证有足够的资源被提供和/或预留，从而足以支持域提供的 SLA。

2.1.1 DS 边界节点和内部节点

DS 域由 DS 边界节点和 DS 内部节点组成。DS 边界节点连接本 DS 域和其它 DS 域或者无 DS 能力的域，DS 内部节点连接同一 DS 域的其它 DS 内部节点或者边界节点。

无论是 DS 边界节点还是内部节点都必须能够按照 DS 编码点信息采用合适的 PHB 转发数据包；否则会导致有不可预测的行为发生。另外，DS 边界节点可能还需要实现其所在 DS 域和其连接的对等 DS 域之间的业务量调节协议（TCA）所定义的业务量调节功能（参见 2.3.3 节）。

内部节点可能会实现有限的业务量调节功能，例如 DS 编码点的重新标记。那些实现了更为复杂的分类和业务量调节功能的内部节点与 DS 边界节点类似（参见 2.3.4.4 节）。

一台 DS 域网络中的主机对于源于其上运行的应用程序的业务流，相当于一个 DS 边界节点；因此我们称这台主机在 DS 域内。如果这台主机并未实现边界节点功能，那么在拓扑结构上最靠近此主机的 DS 节点，将为主机业务流提供 DS 边界节点功能。

2.1.2 DS 入口节点和出口节点

DS 边界节点对于不同方向的业务流，既可以是 DS 入口节点，又可以是 DS 出口节点。业务流在 DS 入口节点处进入 DS 域，在 DS 出口节点处离开 DS 域。DS 入口节点负责保证进入 DS 域的业务流符合本域和此节点直连的另一个域之间的 TCA。DS 出口节点依据两个域之间的 TCA 细节，对转发到其直连的对等域的业务流执行业务量调节功能。注意 DS 边界节点在某些接口中可以作为 DS 内部节点。

2.2 分类业务区域

一个或多个邻接的 DS 域统称为分类业务区域（ DS 区）。DS 区可以支持贯穿区内多个 DS 域的分类业务。

DS 区中的 DS 域可能支持不同的 PHB 组，和编码点到 PHB 的映射规则。不过，为了提供贯穿多个 DS 域的业务，每个对等的 DS 域都必须建立定义（无论显式的或是隐式的）了 TCA 的对等 SLA。TCA 指明了如何在域边界处调节从一个 DS 域传向另一个 DS 域的业务流。

DS 区内的 DS 域也可以采用相同的服务提供策略，并支持相同的 PHB 组和编码点映射。这样的好处是消除了在 DS 域间进行业务量调节的需求。

2.3 业务量分类和调节

分类业务通过在上游网络和下游 DS 域之间建立服务水平协议（ SLA）跨越 DS 域边界。SLA 指定了数据包分类和重标记规则，也指定了业务量简档和对于符合或不符此简档的业务流采取的处理方法（参见 2.3.2 节）。域间的 TCA 就是从 SLA 以直接或间接的方式取得的。

数据包分类策略负责识别出业务量子集，这个子集通过被调节和 /或映射到一个或多个行为集合（通过 DS 编码点重标记）而取得分类服务。

业务量调节包括测量，整形，监察和 /或重标记。其目的是为保证进入 DS 域的业务流符合 TCA 指定的规则。业务量调节的外延依赖于具体的服务细节，涵盖的范围从简单的编码点重标记到复杂的业务监察和整形操作。业务量调节策略的细节应该由网络间协商确定，这个问题不在本文档论述范围内。

2.3.1 分类器

数据包分类器根据数据包包头的某些字段内容选取业务流中的数据包。我们定义了两种分类器。行为集合分类器（ BA 分类器）仅根据 DS 编码点对数据包分类。多字段分类器（ MF 分类器）根据包头中的一个或多个字段值，例如源地址，目的地址， DS 段，协议标识符，源端口号，目的端口号，以及其它信息如引入接口，对数据包分类。

分类器的任务就是选出匹配某种规则的数据包，然后指导它们进入其它的业务量调节器模块接受进一步处理。分类器必须由某个管理例程根据合适的 TCA 进行配置。

分类器还必须鉴别它用来分类数据包的信息的有效性。（ 参见第 6 节）

注意，在上游数据包分片的情况下， MF 分类器在检察传输层包头时，可能将来自同一数据包的后续分片错误分类。这个问题的一种可能的解决方案是保存分片状态信息；然而，由于上游分片可能乱序到达，也可能采取不同的路由，导致这种解决方案缺少一般性。解决数据包分片问题的策略不在本文档论述范围内。

2.3.2 业务量简档

业务量简档描述了分类器选出的业务流的时间特征。它提供了判断一个特定的数据包是否符合业务量简档的规则。例如，一份基于令牌桶的简档可能会如此描述：

codepoint=X, use token-bucket r,b

上面的简档说明，所有 DS 编码点值为 X 的数据包应该通过速率为 r，桶大小为 b 的令牌桶测量器的检测。在本例中，不符合简档的数据包是那些当它们到达时，桶中剩下的令牌已不足的。符合及不符简档这样的两级标准可以扩展到多级。就是说，可以定义多个级别的简档一致性，而不仅是符合，不符合这样两种情况。

对于符合简档和不符合简档的数据包可以采取不同的调节行为，或者不同计费方法。符合简档的数据包无须进一步的调节便可进入 DS 域；或者，可选的，可以改变它们的 DS 编码点。后一种情况发生在 DS 编码点第一次被设为非缺省值时 [DSFIELD]，或者发生在数据包进入一个对此业务流使用不同的 PHB 组或编码点到 PHB 映射策略的 DS 域时。不符合简档的数据包被放入队列，直到它们符合简档（整形），被丢弃（监察），标记一个新编码点（重标记），或者直接转发但需采用另外的计费标准。不符合简档的数据包可能被映射到一个或多个更低优先级的行为集合。这里的更低优先级是指在转发性能的某些方面，低于同类数据包中符合简档的那些所属的行为集合（ BA）。

注意，业务量简档是 TCA 的可选组件，其使用依赖于服务提供和域服务提供策略的详细说明。

2.3.3 业务量调节器

业务量调节器包括下列组件：测量器，标记器，整形器，和丢包器。业务流首先经过分类器的选择，然后被分类器送往业务量调节器的某个组件处。测量器负责（在适当处）测量业务流是否符合业务量简档。测量器对一个特定数据包的测量结果（例如，是否符合简档）会影响对此数据包的标记，丢弃，或整形行为。

当数据包在 DS 边界节点处离开业务量调节器时，每个数据包的 DS 编码点都会被赋予一个适当值。

图 1 说明了分类器和业务量调节器的模块结构。注意，业务量调节器并不一定需要所有四个组件。例如，在没有有效的业务量简档时，数据包可能只通过分类器和标记器。

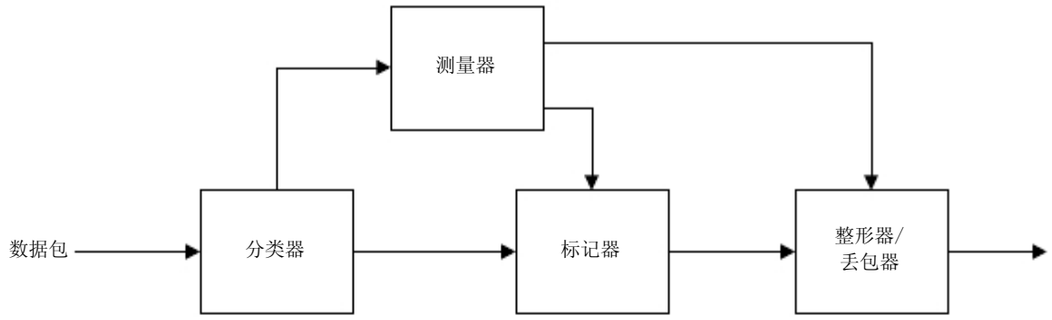

图 1：数据包分类器和业务量调节器逻辑框图

2.3.3.1 测量器

业务量测量器负责测量由分类器根据 TCA 指定的业务量简档选出的数据包流的时间特征。测量器将其测量结果（也称为测量器状态）传递给其它调节功能模块，从而引发对符合或不符（在某种程度上）业务量简档的每个数据包的特殊处理。

2.3.3.2 标记器

数据包标记器负责把数据包的 DS 段设置为特定的编码点值，并将标记过的数据包加入到特定的 DS 行为集合中去。标记器可能被配置为把所有送给它的数据包标记为唯一的编码点值，也可能被配置为根据测量器状态把数据包标记为一些编码点值中的一个值。如果标记器改变了数据包的编码点，那么我们就说标记器“重标记”了此数据包。

2.3.3.3 整形器

整形器负责延迟一个业务流中部分或全部数据包的传输，以便使业务流符合业务量简档的要求。整形器通常有一个有限大小的缓冲区，当缓冲区没有更多的空间容纳需延迟的数据包时，数据包就会被丢弃。

2.3.3.4 丢包器

丢包器负责丢弃一个业务流中部分或全部的数据包，以便使业务流符合业务量简档的要求。这一过程也被称做“监察”业务流。注意，丢包器可以作为一个特殊的整形器（该整形器缓冲区大小为零或仅能容纳几个数据包）而实现。

2.3.4 业务量调节器和 MF 分类器的位置

业务量调节器通常位于 DS 入口和出口边界节点处，但也可能位于 DS 域，或非 DS 域的内部节点处。

2.3.4.1 在源域内

我们定义源域为发起接受特殊服务的业务流的节点所在的 DS 域。位于源域中的业务源和媒介节点可以实现业务量分类和调节功能。从源域中发出并穿越边界的业务流可能直接被业务源做上标记，或者在离开源域之前由媒介节点标记。这两种方式分别被称为“初始标记”和“预标记”。

考虑这样一个例子：在一家公司中， CEO 的数据包通常要求有较高优先级。 CEO 的主机会把所有其发出的数据包的 DS 编码点标记为一个代表“较高优先级”的值。或者，由 CEO 主机直接连接的第一跳路由器负责把 CEO 的数据包分类，并做适当的标记。象这样的高优先级业务流也可能在靠近数据源处进行调节，以便对特定数据源发出的高优先级业务的总量有所限制。

在业务源处对数据包进行标记有几点优势。首先，业务源更容易获得应用程序的需求。因此，它在确定哪些数据包应该享受更好的转发待遇时，可以将应用程序的需求纳入考虑。另外，在业务流与来自其它数据源的业务流合并之前对其数据包分类，要更简单。因为此时一个节点所使用的分类规则的数量会较少。

鉴于数据包的标记可能分散在多个节点处进行，源 DS 域有责任保证流向其服务提供者 DS 域的业务流集合与适当的 TCA 相符合。额外的分配机制，如带宽代理或 RSVP，可能被用来为提供者网络中特定的 DS 行为集合动态分配资源 [3BIT，Bernet] 。源域的边界节点应该保证业务流符合 TCA，必要时，要对数据包监察，整形，或重标记。

2.3.4.2 在 DS 域边界

业务流可能在边界连接的任何一端（上游域 DS 出口节点或者下游域 DS 入口节点）被分类，标记或者调节。域间的 SLA 应指明由哪个域负责将业务流映射到 DS 行为集合，以及调节这些集合使之符合适当的 TCA。然而， DS 入口节点必须假定流入的业务流不符合 TCA，因此必须准备根据本地策略强制执行 TCA。

如果数据包在上游域中被预标记和调节，那将意味着下游域只需支持很少的分类和业务量调节规则。在这种情况下，下游 DS 域可能只需要根据 TCA 对流入的行为集合重标记或监察。然而，那些具有路径依赖或源依赖性的更复杂业务可能还需要下游 DS 域入口节点进行 MF 分类。

如果 DS 入口节点与一个无 DS 功能的上游域连接，那么 DS 入口节点就必须能对流入的业务执行所有需要的业务调节功能。

2.3.4.3 在无 DS 功能的域内

在无 DS 功能的域内的业务源或媒介节点可以使用业务量调节器在业务流到达下游 DS 域入口节点之前预标记之。这样，本地分类和标记策略将被隐藏。

2.3.4.4 在内部 DS 节点处

尽管基本体系结构假设复杂的分类和业务量调节功能位于网络的入口和出口边界节点处，在网络内部节点处配置这些功能也并未被排除。例如，在一条越洋链路上，需要有更多更严格

的接入策略，这就需要在这条链路的上游节点处实现 MF 分类和业务量调节功能。当然，这种方法在可扩展性上有些限制。因为那将意味着在一个节点上，维护大量的分类和调节规则。

2.4 每一跳行为

每一跳行为（ PHB）是指 DS 节点运用于特定 DS 行为集合上的，外部可观察的转发行为。“转发行为”在这里是一个广义概念。例如，当仅有一个行为集合占用一条链路时，可观察的转发行为（如，丢包率，延迟，时延抖动）就只依赖于链路的相对负载（即是说，在“行为”采用一种工作保存式的调度策略）。有意义的行为上的差别通常产生于在同一个节点，多个行为集合竞争缓冲区和带宽资源的情况下。 PHB 是节点给行为集合分配资源的一种方法，正是基于这种逐跳进行资源分配的机制，我们才构筑了分类业务模型。

PHB 的最简单例子是保证至少把一条链路带宽的 X%（在一定的时间间隔内）分配给一个行为集合。这种 PHB 在各种业务竞争条件下都可以被公正并且很容易的测量。另一种稍复杂点的 PHB 要求确保最少占有 X%的链路带宽，同时享受相应份额的链路剩余带宽。一般来说，PHB 的可观察行为依赖于对相关行为集合或其它行为集合的业务量特性的约束。

PHB 通过指定其相对于其它 PHB 的资源（如，缓冲区，带宽）优先级来定义，也可能通过它们的可观察业务量特性（如，延迟，丢包率）来定义。这些 PHB 可以作为资源分配的基石，并且一致性起见，应被指定为一组（ PHB 组）。PHB 组中的每一 PHB 都享有共同的限制，例如数据包安排或者缓冲区管理策略等。同组的 PHB 间的联系在于它们绝对的或者相对的优先级（例如，采用确定阈值或随机阈值的丢包优先级），但是这并不是必须的（例如， N 等分链路资源）。一个单独定义的 PHB 可以看作是 PHB 组的特例。

在节点处， PHB 是通过一定的缓冲区管理和数据包安排策略实现的。 PHB 是通过与服务提供策略相联系的行为特征定义的，而不是根据采取了何种实现机制。一般来说，可以有很多种实现机制去实现特定的 PHB 组。而且，在一个节点上，可以实现多于一个的 PHB 组，并在域内使用。所定义的 PHB 组应该保证适当的组间资源分配简单易行，并且能够实现同时支持两组或更多组的集成机制。一个 PHB 组定义时，应指明其与已有组之间可能的冲突。这些冲突可能来自于有些操作不允许同时执行。

如[DSFIELD] 中描述，在节点处，根据收到数据包的 DS 编码点选择 PHB。标准化的 PHB 有推荐的编码点。然而，全部编码点空间远大于分配给标准化 PHB 使用的编码点空间，[DSFIELD] 把剩余空间提供给了局部使用。编码点到 PHB 的映射表可以即包括一对一，也包括 N 对一的映射。注意，所有的编码点都必须被映射到某一 PHB：在缺少某些局部策略的情况下，那些没有映射到标准化 PHB 的编码点应该被统一映射到一个缺省 PHB。

2.5 网络资源分配

在 DS 域节点上实现，配置，操作和管理的 PHB 组，应能根据域服务提供策略，有效的分配使用这些节点的资源，以及节点间链路。业务量调节器可以通过执行 TCA，或者从域中节点或其它业务量调节器取得反馈，从而更有效的控制资源的使用。尽管在没有复杂的业务量调节功能时，也可以提供很多服务（例如，仅使用静态标记策略），但类似于监察，整形，和动态重标记这样的功能，可以允许向用户提供具有量化的性能参数的服务。

业务量调节器及内部节点间的配置和交互需要有域高层的管理控制，可能还需要一个控制实体和适当的协议。控制模型的实现方案有很多种。

这些模块之间交互的准确特征和实现细节超出了本体系结构的范围。然而，可扩展性要求域的控制不需要网络资源的微管理。最具扩展性的控制模型应在开环方式下在操作时隙内操作

节点，并且由于 SLA 是变化的，所以只需要管理时间刻度内的管理操作。这种简单模型可能在某些情况下并不适用，此时，一些自动的但缓慢改变的操作控制（按分钟而不是秒）在平衡对网络资源的适用方面就会更具吸引力。

3 每一跳行为（ PHB）的规范设计指导方针

对每一跳行为进行标准化的基本要求在 [DSFIELD] 中给出。本节详细阐述 PHB（组）定义时的其它要求。主要目的是帮助建立 PHB 实现时的一致性。当一个 PHB 组标准化时，它必须满足这些要求，从而保持本体系结构的完整性。

G.1：一个标准 PHB 必须从为标准映射保留的编码区域内 [DSFIELD] ，选择一个推荐的 DS 编码点。推荐的编码点由 IANA 指定。一个 PHB 提议可以从 EXP/LU 空间内选取一个临时编码点，以便进行域间实验。注意，不能要求检查包头除 DS 域以外的信息域来确定该数据包所对应的 PHB 。

G.2：每一个新提出的 PHB 组规范应该包括其行为及行为目的的概述。概述中应包括此 PHB 组要解决的问题的描述。还应包括此 PHB 组涉及的基本概念。这些概念应包括，但不仅限于，队列行为，丢弃行为，和输出链路选择行为。最后，概述中应阐明此 PHB 组解决所描述问题的方法。

G.3：PHB 组规范应说明其包括的单个 PHB 的个数。当 PHB 组包括多于一个 PHB 时，PHB 组规范就必须对这些 PHB 间的交互和限制做出详细说明。例如，规范必须说明如果同一微流中不同的数据包被标记为此组中不同的 PHB 时，数据包被重新排序的可能性是否会增大。

G.4：如果 PHB 组的特定功能依赖于某些外部限制，比如服务提供限制，那么 PHB 定义中，就必须说明当这些限制被违反时，此 PHB 所产生的行为。更进一步，如果这些限制被违反时，又需要采取类似包丢弃或重标记的行为，那么这些行为应被特别规定。

G.5： 一个 PHB 组可能被指明为仅在一个域内局部使用。其作用是在域内提供面向特定域的功能或服务。在这种情况下， PHB 规范有助于向服务提供者提供一份关于此 PHB 组的一致性定义。然而，任何只定义为局部使用的 PHB 组，不能进行标准化。但是，它们可以作为知识性 RFC 文档发表。相反的，作为通用的 PHB 组，就必须经过严格的标准化过程。所以，所有的 PHB 组提议都必须明确指明其用途是通用的还是局部的。

PHB 组可能被设计来提供主机到主机， WAN 边界到 WAN 边界，和 /或域边界到域边界服务。出于一致性考虑，在 PHB 定义中使用“端到端（ host to host ）”应被理解为“主机到主机”。

由于实验或操作要求，在域内还可以定义和配置其它 PHB 组。这些局部有效的 PHB 组并不需要公开定义，但是它们应使用 EXP/LU 空间的编码点（ [DSFIELD] 中定义）。

G.6：被标记为 PHB 组中某个 PHB 的数据包，有可能在域中或域边界处，重标记为同 PHB 组中的另一个 PHB 。有三种典型情况需要这种转换：

a. 与此 PHB 组对应的编码点，也被用来携带网络状态信息。

b. 需要提升或降低数据包的 PHB 的情况。（如果组内的 PHB 可进行某种排序）

c. 域边界处缺少 SLA 。这种情况下，数据包通过边界链路时的编码点 /PHB 根据上游域的局部策略决定。

PHB 规范应该清楚的说明数据包被重标记（例如，提升或降低）为本组中另一个 PHB 的情况。如果某些情况下不能改变数据包的 PHB，那么规范中必须清楚说明此时改变 PHB 的后

果。改变数据包的 PHB（无论改为同组的，还是不同组），可能引起的后果之一是，对微流中数据包重新排序的可能性增加。 PHB 可能携带某些主机到主机， WAN 边界到 WAN 边界，和/或域边界到域边界的语义，这些语义很难在数据包重标记为组内（或其它组）另一个 PHB 时保存下来。

对于某些 PHB 组，可以通过重标记数据包为组内其它 PHB 来反映节点状态的变化。如果一个 PHB 组被设计来反映网络状态，那么在相应的 PHB 规范中，就必须详述组中各个 PHB 与其反映的网络状态之间的联系。此外，如果这些 PHB 对节点转发行为有所限制，那么这些限制可以被表达为节点应当，或必须执行的行为。

G.7：PHB 组规范中应包括一节说明隧道技术在本组的应用。此节应说明当数据包被封装入隧道后，如何根据其内层包头的 DS 域，来确定外层包头 DS 域。还应说明，在隧道出口处，内层包头应做如何的改变。（参见 6.2 节）。

G.8：定义 PHB 组是一个增量式的过程。当定义新的 PHB 组时，应说明它如何与已定义的 PHB 组协调。每一个新建立的 PHB 组，其作用的范围可能是全新的，也可能是已有 PHB 组的扩展。如果一个 PHB 组完全独立于某些或全部已有 PHB 组规范，那么在其规范中应有一节详述它如何与已经标准化的 PHB 组共存。例如，这一节中应该说明当微流中的数据包被标记为与两个不同 PHB 组都相关的编码点时，它们被重排序的可能性有多大。如果在同一节点中同时操作两个（或更多） PHB 组不可能或是有害的，这种情况必须在此节中说明。如果节点在处理同时收到的被标记为两个（或更多） PHB 组的多个数据包时，需采取某些特殊动作，那么这些特殊动作也应在此节说明。

定义 PHB 组时，应注意避免循环定义。

如果提议中的 PHB 组是已有的某个 PHB 组的扩展，那么规范中应包括一节详述此 PHB 组如何与被扩展 PHB 组交互。更进一步，如果此 PHB 组改变或更狭义的定义了某些已有行为，这些情况也应在规范中说明。

G.9：每个 PHB 规范都应包括一节阐述实现此 PHB 组的最小一致性需求。这一节主要目的是提供此 PHB 组实现时容许变化的行为的变动范围。这一节可以采用规则，表格，伪代码，或测试的形式叙述。

G.10：PHB 规范应包括一节详述其行为的安全性。此节应包括对在隧道出口改变内层包头的编码点，以及这种改变对数据包转发行为的影响的讨论。

另外，此节还应讨论提议的 PHB 组如何被用在拒绝服务攻击，减少服务协议攻击，和违反服务协议攻击。最后，此节应讨论检测这些攻击的可能方案。

G.11：PHB 规范中应包括一节详述配置和管理问题。这些问题可能影响 PHB 的操作和使用此 PHB 的候选服务。

G.12：强烈建议每个 PHB 规范都提供一个附录，详述其对现有服务和潜在服务提供的服务行为特征。这些服务包括但不限于用户特定的，设备特定的，域特定的或端到端的服务。强烈建议附录中包括一节描述用户，设备，和 /或域如何验证这些服务。

G.13：建议每个 PHB 规范提供一个附录。此附录面向域内部使用，提供本域向不支持此 PHB 组的对等域转发数据包时选择何种 PHB 的指导。

G.14：建议每个 PHB 规范提供附录说明提议的 PHB 组对已有的高层协议的影响。在某些情况下， PHB 允许对高层协议做可能的改变，这些改变可能增加，也可能减少提议的 PHB 组的使用。

G.15：建议每个 PHB 规范提供附录，对使用共享媒体或交换的链路层提供从 PHB 到链路层 QoS 机制映射的建议。在 PHB 和链路层 QoS 机制之间进行适当的映射涉及许多因素，这个问题已超出本文档的论述范围；然而，在 PHB 规范中应设法给出一些指导。

4 与非分类业务兼容节点的互操作

我们把非分类业务兼容节点（ non-DS-compliant node ）定义为任何不会根据 [DSFIELD] 描述译解 DS 域，和/或没有实现部分或全部标准化 PHB 组（或只在特定 DS 域使用的 PHB 组）的节点。这可能域节点的能力或配置有关。我们把遗留节点（ legacy node ）定义为非分类业务兼容节点的特殊情况，这些节点实现了 IPv4 优先级和 [RFC791，RFC1812] 定义的转发策略，但在其它方面是非分类业务兼容的。在 [DSFIELD] 中定义的等级选择编码点（Class Selector Codepoints ）被特别考虑，保持与 IPv4 中 TOS 字节定义的优先值兼容，同时，在 [DSFIELD] 中定义的等级选择 PHB 也与 [RFC791，RFC1812] 定义的优先转发行为相兼容。遗留节点和分类业务兼容节点之间的关键区别在于，遗留节点可能，也可能不译解 [RFC1349] 定义的 TOS 字节的 3-6 比特（“DTRC”比特）；实际上它不会译解 [DSFIELD] 指定的那些比特。我们假设使用 [RFC1349] 中定义的 TOS 标记是被反对的。注意，那些不是遗留节点的非分类业务兼容节点，在转发带有非零 DS 编码点的数据包时，其转发行为是不可预测的。

分类业务依赖于基于在节点上实现每一跳行为的资源分配机制。如果业务流穿过非分类业务兼容节点或无分类业务能力域，那么服务的质量或统计确保水平很可能会下降。

我们讨论两种情况。第一种情况关于在 DS 域中使用非分类业务兼容节点。注意， PHB 转发主要是在对稀有节点和链路资源进行有控制的分配时有用。在高速，轻载荷链路上，最坏包延迟，抖动，和丢失都是可以忽略的，并且，在这样的链路上游使用非分类业务兼容节点不会造成服务降级。在多数实际情况下，一个节点缺少 PHB 转发支持会导致无法对通过此节点的数据流提供低时延，低丢包率服务，或预定服务带宽。然而，如果在 DS 域中能够仅使用 [DSFIELD] 定义的等级服务编码点，并且在遗留节点上的优先级实现方案与路径上的其它节点的转发行为兼容，那么在无法兼容分类业务的节点处使用遗留节点就会是一个较好的替代方案。注意，由于遗留节点可能，也可能不会根据 [RFC1394] 译解第 3-5 比特，从而导致不可预测的转发行为，所以，必须限制只使用等级选择编码点。

第二种情况关于穿越无 DS 能力域（ non-DS-capable domain ）的服务行为。出于讨论目的，我们假定无 DS 能力域不会在域边界节点处进行业务量调节；因此，即便在域内存在遗留节点或 DS 兼容节点的情况下，由于在边界处缺少业务量管理，也会限制本域，使之不能稳定提供某些服务。 DS 域和无 DS 能力域之间可以就如何在进入无 DS 能力域之前在 DS 域出口处标记数据包达成协议。这个协议可通过业务量采样而不是严格的业务量调节确保实施。 如果已知无 DS 能力域中包含遗留节点，那么上游 DS 域可以选择将分类业务流重标记为一个或多个等级选择编码点。如果对下游域的业务量管理能力一无所知，而且缺少域间协议，那么 DS 出口节点可以选择将 DS 编码点重标记为零，并假设无 DS 能力域会平等对待所有业务流，遵从尽力而为原则转发。

在无 DS 能力域与 DS 域对等的情况下，来自无 DS 能力域的业务流应在 DS 域入口节点处根据适当的 SLA 或其它协议进行调节。

5 关于组播

对组播业务使用分类业务机制引入很多问题。首先，在某个 DS 域入口节点进入的组播数据包可能会被复制，然后通过多条不同的路径穿越此 DS 域。这种情况下，组播业务就会比单播业务消耗更多的网络资源。由于组播成员是随时变化的，因此很难预测从上游网络来的某一组播业务会消耗本域的多少网络资源。这种不确定性的后果之一就是很难向组播业务提供者提

供定量服务保证。此外，有必要为单播业务保留编码点和 PHB，以便提供不受组播干扰的资源。

第二个问题是到达 DS 入口点的组播数据包如何选择 DS 编码点。组播数据包可能在多个出口节点流出 DS 域，每个出口节点都对应一个下游 DS 域，和相应的 SLA。应注意组播数据包的编码点不能违反任何一个 SLA。当为从某一入口节点进入的请求分类服务的组播业务流建立分类器和业务量调节器状态时，应该考虑到相应的下游 DS 域标识和 SLA 细节（使之服从路由策略及路由体系的稳定性）。这样相应的下游 DS 域 SLA 可以部分的在上游 DS 域入口点处得以履行，从而减轻上游域入口点的分类和业务量调节负担。当然，在组播业务中采取这样的方法很困难，因为组播成员是动态变化的。结果是单播业务的服务保证会受到影响。解决此问题的一种方案是，为组播业务创建单独的 SLA，并且或者让组播业务采用特殊的一系列编码点，或者在 DS 出口点实现必要的分类和业务量调节机制，根据下游域 SLA 将单播业务优先分离，提供服务。

6 安全和隧道问题

这一节主要讨论由于引入分类业务而引起的安全问题，主要是拒绝服务攻击隐患，和通过未授权业务窃取服务隐患（参见 6.1）。另外，在 IPsec 条件下分类业务的操作和与 IPsec 交互问题也有讨论（参见 6.2 节）。审查需求也在 6.3 节有所讨论。本节讨论的问题包括由 IPsec 和非 IPsec 隧道引入的。

6.1 窃取和拒绝服务

分类业务的主要目的是在统一的网络体系结构上，向业务流提供不同级别服务。很多资源管理技术被用来完成这一目标，最终的结果是有些数据包会受到不同的（例如，更好的）服务。把网络业务映射到特定的行为集合从而取得不同的（或好或坏）服务，主要依靠设定 DS 域。随之而来的问题就是，可以通过把 DS 域修改为代表更高优先级的编码点，或者直接注入 DS 域置为这些编码点的数据包而获得更好的服务。在极端情况下，如果修改或者注入的数据包耗尽了网络资源，这种窃取服务行为就演变成拒绝服务攻击。解决方案包括综合使用在 DS 边界节点进行业务量调节和增强 DS 域内网络结构的安全性和完整性。

如第 2 节描述， DS 入口节点必须调节所有进入 DS 域的业务流，以确保它们带有可接受的 DS 编码点。就是说编码点必须符合适用的 TCA 和域服务提供策略。因此，入口节点是防止基于修改 DS 编码点（例如，改为此业务流无权的 DS 编码点）的窃取和拒绝服务攻击的主要战线。因为，任何这样的攻击都会违反 TCA 和/或服务提供策略。一种重要类型的入口节点是在 DS 域中可以发起业务的节点。必须保证任何发起的业务都带有可接受的 DS 编码点。

域服务提供策略和 TCA 都可能要求入口节点改变某些流入数据包的 DS 编码点（例如，入口路由器可能根据适当的 SLA 设置用户业务流的 DS 编码点）。入口节点必须调节所有的业务流以保证 DS 编码点是可接受的；带有不可接受编码点的数据包必须或者被丢弃，或者将其 DS 编码点在转发前改为可接受值。例如，入口节点可能会将从与之无增强业务协议的域收到的业务流 DS 编码点设置为缺省 PHB 编码点 [DSFIELD] 。某些 DS 编码点的适用可能需要特别的业务授权（例如，那些对应于增强业务的编码点），并且这类授权可以采用技术方法（如， IPsec ）和/或非技术方法（如，固定连接到唯一客户的入口链路）。

域间协议可以使上游域部分或全部保证业务流带有下游域可接受的 DS 编码点，从而降低入口节点进行业务量调节的要求。这种情况下，入口节点仍可以进行业务量调节检查，以避免对上游域的依赖（例如，这类检查可以防止跨越域边界传播的窃取服务攻击）。如果由于上游域未能履行其责任，导致业务流未能通过入口节点检查，那么这一审查事件必须被记录；记录

项包括接收到数据包的日期/时间，源和目的 IP 地址，和导致未通过此检查的 DS 编码点值。实际中，需要在这类检查的好处和它们对系统性能的影响之间做出权衡，确定到底哪些情况需要检查。

DS 域内部节点依赖于 DS 段向要求分类业务功能的业务流提供服务。节点必须依靠正确的 DS 域操作来防止带有不可接受 DS 编码点的业务流到达。健壮性考虑要求带有不可接受 DS 编码点数据包的到达不会引起网络节点瘫痪。内部节点没有义务监督服务提供策略 SLA 的执行情况，也不需要对 DS 编码点进行检查。内部节点可以对 DS 编码点执行某些业务量调节检查（例如，检查出在某条特定链路上从不会使用的编码点），从而提高节点安全性和健壮性（例如，可以防止基于 DS 编码点修改的窃取服务攻击）。任何检查出的错误都是审查事件，而应被记录。记录项包括接收到数据包的日期/时间，源和目的 IP 地址，和导致未通过此检查的 DS 编码点值。实际中，需要在这类检查的好处和它们对系统性能的影响之间做出权衡，确定内部节点处到底需做哪些检查。

如果一条链路不能充分保证其安全性，即是说，不能防止 DS 编码点修改或者注入不合法业务流，那么就应该被当作边界连接对待（即，任何从此链路到达的业务流都象从入口节点处进入 DS 域的业务流那样进行各种检查）。局部安全策略提供关于“充分安全”的定义。“充分安全”定义中应给出一个 DS 编码点修改和/或业务流注入的危险和后果，与对链路的采取额外安全措施的平衡点。链路安全性可以通过物理接入控制和/或软件方法，如确保数据包完整性的隧道技术，得到提高。

6.2 IPsec 和隧道交互

在 IPsec 协议中，如 [ESP，AH] 描述，不含有对 IP 头 DS 段任何形式的加密（在隧道中，外层 IP 头的 DS 段未被加密）。因此，网络节点对 DS 段的修改不会影响 IPsec 端到端的安全性，因为这种修改不会引起 IPsec 完整性检查失败。其结果是，IPsec 不能提供任何措施防止对 DS 段的修改（如，中间人攻击 man-in-the-middle attach），因为攻击者的修改不会破坏 IPsec 的端到端完整性。在有些环境，这种修改 DS 段而不影响 IPsec 完整性检查的能力，可以组成一个隐蔽通道；如果有必要消除这样的通道或减少其带宽，可以把 DS 域配置为，可以在业务流离开更高安全性域的 DS 出口节点处进行所需的处理（如，将敏感业务流的所有 DS 段置为同一个值）。

IPsec 的隧道模式为封装的 IP 头 DS 段提供了安全性支持。隧道模式下的 IPsec 包含有两个 IP 头：由隧道入口节点提供的外层包头和由数据包源节点提供的被封装的内层包头。当一条 IPsec 隧道（部分或全部的）经过分类业务网络时，中间的网络节点会对外层包头中的 DS 段进行操作。在隧道出口节点处，IPsec 会去掉外层包头，并（如果需要）使用内层包头转发数据包。如果内层 IP 头没有被隧道出口节点所在 DS 域的某个 DS 入口节点处理过，那么隧道出口节点就作为离开隧道的业务流进入 DS 域的入口节点，也因此，此节点必须履行相应的业务量调节策略（参见 6.1 节）。如果 IPsec 对所封装数据包有足够强壮的密码完整性检查（这里的“足够”取决于局部安全策略），那么隧道出口节点就可以安全的假设内层包头的 DS 段值与在隧道入口时相同。这就允许与隧道入口节点位于同一 DS 域的隧道出口节点，可以象对待从同一 DS 域其它节点来的数据包一样（即是说，省略 DS 入口节点业务量调节处理）处理从隧道流出的数据包，并保证安全性。这样做的一个重要后果就是，局部于 DS 域的其它不安全链路可以使用足够强壮的 IPsec 隧道使其安全。

此分析和隐含适用于任何进行完整性检查的隧道协议，只是对内层包头 DS 段的确保水平依赖于隧道协议进行的完整性检查力度。在隧道可能穿越当前 DS 域之外的节点时，如果缺少对这类隧道足够的信任，封装的数据包就必须象对待从域外到达 DS 入口节点的数据包同样处理。

当前 IPsec 协议不允许在隧道出口节点处除 IPsec 封装时改变内层包头的 DS 段。这保证了不能利用修改 DS 段的方法穿越 IPsec 隧道终点实施窃取服务或拒绝服务攻击，因为任何这类修改都会在隧道终点处丢弃。本文档不对 IPsec 做任何修改。

如果 IPsec 的未来版本允许在隧道出口节点处根据外层包头 DS 段值修改内层包头 DS 段值（例如，复制外层 DS 段的部分或全部内容到内层 DS 段），那么就还需对由此而来的安全性问题额外考虑。当一条隧道完全处于一个 DS 域中，并且这条链路绝对安全，外层 DS 段不会被修改，在修改内层 DS 段时的唯一限制来自于域服务提供策略。另外，进行这种修改的隧道出口节点对流出隧道的业务流来说，应是 DS 入口节点，必须实施必要的业务量调节功能，包括防止窃取服务和拒绝服务攻击（参见 6.1 节）。如果隧道不是在其出口节点处进入 DS 域，那么隧道出口节点就依靠上游的 DS 入口节点确保外层 DS 段值是可接受的。即使在这种情况下，也有某些检查是必须由隧道出口节点实施的（例如，加密隧道内层和外层 DS 段值一致性检查）。任何检查出的错误都必须留有审查记录，记录内容包括数据包到达的日期/时间，源和目的 IP 地址，和所携带的不可接收 DS 编码点。

从体系结构角度，至少可以有两种不同的方式看待 IPsec 隧道。如果隧道被看作逻辑上只有一跳的“虚链路”，那么隧道中间节点转发隧道中数据流的行为，对隧道终点来说就应该是不可见的，并且在解除封装过程中也不应该修改 DS 段值。相反的，如果隧道被看作是多跳的，那么在隧道解除封装的过程中修改 DS 段值，就是可以的了。后一种情况的具体例子如下。隧道终止于 DS 域内部节点，而域管理者并不想在此节点安装业务量调节功能（例如，为了简化业务量管理）。一种解决方案是，根据外层 IP 包头的 DS 编码点（根据此值在 DS 入口节点处进行业务量调节）设置内层 IP 包头 DS 编码点。这样就有效将业务量调节功能从 IPsec 隧道出口节点转移到了适当的上游 DS 入口节点（ DS 入口节点必须已经对未解除封装的业务流进行了业务量调节）。]

6.3 审查

并不是所有支持分类业务的系统都要实现审查功能。然而，如果分类业务支持融入了一个提供审查功能的系统中，那么分类业务实现也必须支持审查。如果支持审查功能，那么就必须允许系统管理者整体性的开启或禁止分类业务审查功能，并且允许部分的开启或禁止这类审查。

大多数情况下，审查功能的粒度是局部问题。然而，本文档中定义了一些审查事件，以及每一事件记录应包括的最小信息集。额外信息（如，引起此审查事件的数据包本身）也可被包括在记录信息中。并未在本文档中提到的其他事件，也可以导致一条审查记录。检测到审查事件后，并不要求接收者向发送者传送任何的消息，因为回传任何消息都有可能导致包括拒绝服务攻击在内的危险。

7 感谢

本文档得益于早期由 Steven Blake, David Clark, Ed Ellesson, Paul Ferguson, Juha Heinanen, Van Jacobson, Kalevi Kilkki, Kathleen Nichols, Walter Weiss, John Wroclawski, 和 Lixia Zhang 撰写的初稿。

很多人为本文档提供了有帮助的建议和意见，作者在此向他们表示感谢： Kathleen Nichols, Brian Carpenter, Konstantinos Dovrolis, Shivkumar Kalyana, Wu-chang Feng, Marty Borden, Yoram Bernet, Ronald Bonica, James Binder, Borje Ohlman, Alessio Casati, Scott Brim, Curtis Villamizar, Hamid Ould- Brahi, Andrew Smith, John Renwick, Werner Almesberger, Alan O'Neill, James Fu, 和 Bob Braden.

8 参考文献

[802.1p] ISO/IEC Final CD 15802-3 Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Common specifications - Part 3: Media Access Control (MAC) bridges, (current draft available as IEEE P802.1D/D15).

[AH] Kent, S. and R. Atkinson, "IP Authentication Header", RFC 2402, November 1998.

[ATM] ATM Traffic Management Specification Version 4.0 <af-tm-0056.000>, ATM Forum, April 1996.

[Bernet] Y. Bernet, R. Yavatkar, P. Ford, F. Baker, L. Zhang, K.Nichols, and M. Speer, "A Framework for Use of RSVP with Diff-serv Networks", Work in Progress.

[DSFIELD] Nichols, K., Blake, S., Baker, F. and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474 , December ， 1998.

[EXPLICIT] D. Clark and W. Fang, "Explicit Allocation of Best Effort Packet Delivery Service", IEEE/ACM Trans. on Networking, vol. 6, no. 4, August 1998, pp. 362-373.

[ESP] Kent, S. and R. Atkinson, "IP Encapsulating Security Payload (ESP)", RFC 2406 , November 1998.

[FRELAY] ANSI T1S1, "DSSI Core Aspects of Frame Rely", March 1990.

[RFC791] Postel, J., Editor, "Internet Protocol", STD 5, RFC 791 , September 1981.

[RFC1349] Almquist, P., "Type of Service in the Internet Protocol Suite", RFC 1349 , July 1992.

[RFC1633] Braden, R., Clark, D. and S. Shenker, "Integrated Services in the Internet Architecture: An Overview", RFC 1633, July 1994.

[RFC1812] Baker, F., Editor, "Requirements for IP Version 4 Routers", RFC 1812 , June 1995.

[RSVP] Braden, B., Zhang, L., Berson S., Herzog, S. and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205 , September 1997.

[2BIT] K. Nichols, V. Jacobson, and L. Zhang, "A Two-bit Differentiated Services Architecture for the Internet", ftp://ftp.ee.lbl.gov/papers/dsarch.pdf , November 1997.

[TR] ISO/IEC 8802-5 Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Common specifications - Part 5: Token Ring Access Method and Physical Layer Specifications, (also ANSI/IEEE Std 802.5- 1995), 1995.

9 作者联系地址

Steven Blake

Torrent Networking Technologies

3000 Aerial Center, Suite 140

Morrisville, NC 27560

Phone: +1-919-468-8466 x232

EMail: slblake@torrentnet.com

David L. Black

EMC Corporation

35 Parkwood Drive

Hopkinton, MA 01748

Phone: +1-508-435-1000 x76140

EMail: black_david@emc.com

Mark A. Carlson

Sun Microsystems, Inc.

2990 Center Green Court South

Boulder, CO 80301

Phone: +1-303-448-0048 x115

EMail: mark.carlson@sun.com

Elwyn Davies

Nortel UK

London Road

Harlow, Essex CM17 9NA, UK

Phone: +44-1279-405498

EMail: elwynd@nortel.co.uk

Zheng Wang

Bell Labs Lucent Technologies

101 Crawfords Corner Road

Holmdel, NJ 07733

EMail: zhwang@bell-labs.com

Walter Weiss

Lucent Technologies

300 Baker Avenue, Suite 100

Concord, MA 01742-2168

EMail: wweiss@lucent.com

10 完整版权声明