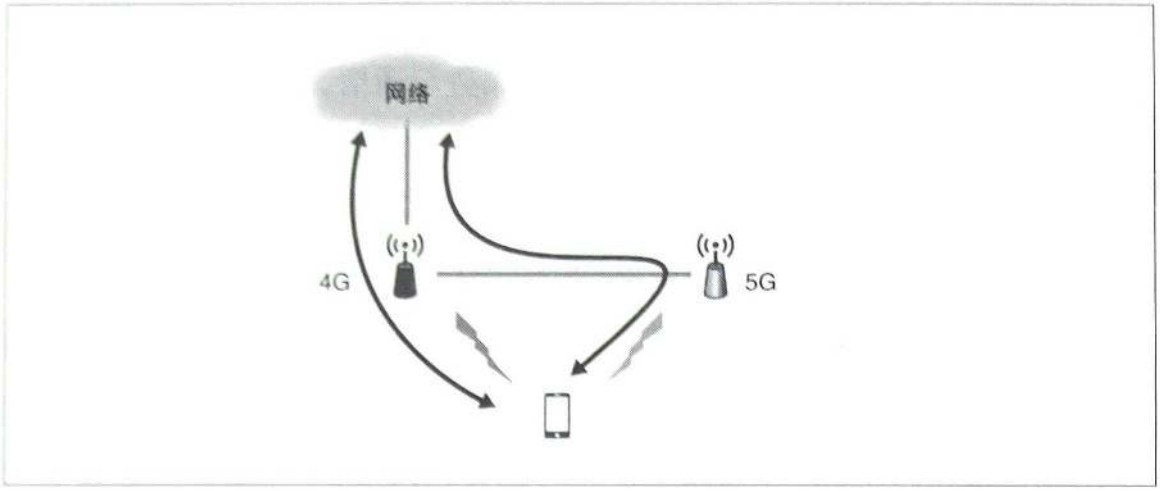


图 7.12 4G 与 5G 构成的双连接

[TCP 相关动向 ❸] 高清流媒体

如上所述，4K、8K 大容量视频的数据传输也是 5G 的一种应用。同时，5G 在投入使用后，想必也会像刚才说的那样，与 4G 的网络一起并行运作。在这种应用场景下，必须采用同时符合 4G 和 5G 的网络容量（传输速度）的数据量进行数据传输。如果是视频，那么由于可以通过编码方法控制画质，所以只要能识别网络的通信方式就没有问题。

在通常情况下，实时视频传输会使用 UDP 协议，然而在传输高清视频时，如果随心所欲地发送大量数据，想必一定会引发拥塞。因此，TFRC（TCP Friendly Rate Control protocol，TCP 友好速率控制协议）这种考虑了与 TCP 的公平性，在 UDP 中增加了拥塞控制功能的协议逐渐被广泛使用。此协议规定在 RFC 3448 中，其基本思路与 TCP 类似，在数据传输开始时使用慢启动，之后在拥塞状态下使用 *RTT* 等控制数据传输量。

然而，在支持大容量通信的 4G、5G 网络中进行实时传输时，缓慢增大窗口大小的协议恐怕会成为瓶颈，这一点令人不安。最近有一项研究，其内容是通过某种外部信息判断出当前网络的通信方式（是 4G 还是 5G 等），然后根据网络容量主动控制数据的传输量，以此最大限度地利用通信资源，实现流畅的流媒体传输。

7.3 物联网

通过互联网控制各种各样的设备

近些年来，通过互联网控制各种各样设备的**物联网**服务逐渐流行开来。本节将介绍物联网的概要、其在通信视角下的问题，以及对应的 TCP 相关动向。

[背景] 多样的设备和通信方式

物联网的总体结构如图 7.13 所示。各种各样的设备接入互联网，并与云服务等服务器相连接。物联网的设备多种多样，其中有直接接入互联网的，也有通过网关接入互联网的。

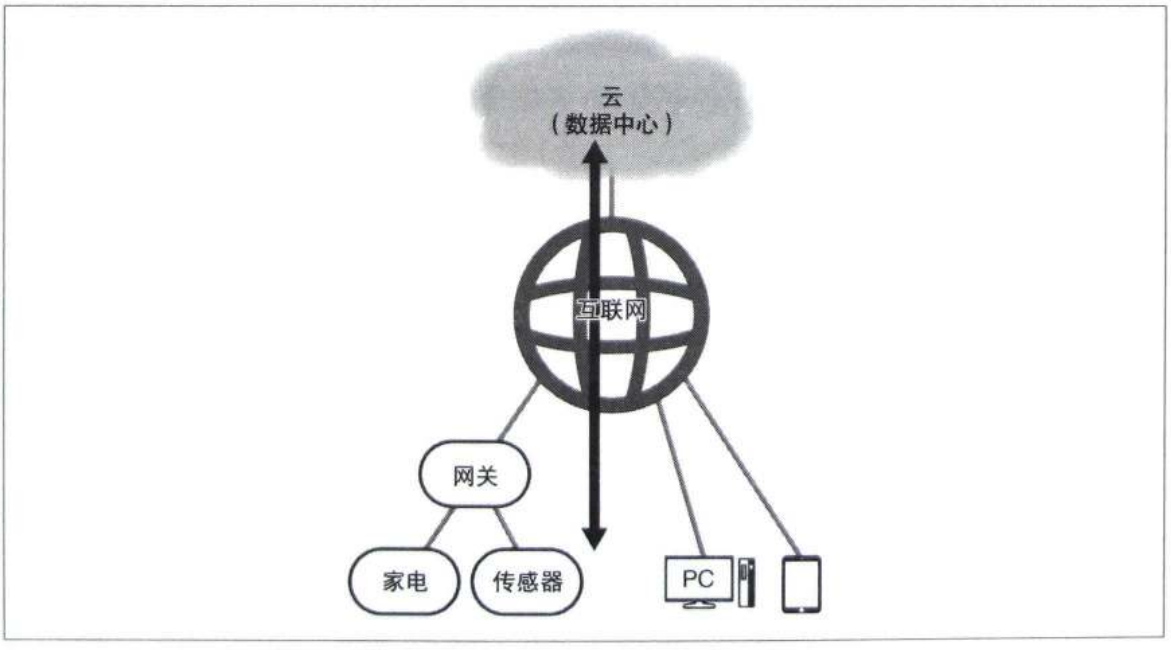


图 7.13 物联网的总体结构

设备收集到的数据将被发送给服务器，在服务器进行数据处理。处理完的数据有时会发送到外部设备，由设备使用，有时则由服务器根据处理结果直接控制设备。

传感器设备的示例

举例来说，人们常见的一种物联网服务就是大楼里面放置的温度传感器会将收集到的温度数据发送给服务器，然后当室内温度超过一定值时，服务器就会打开空调。通过物联网，今后会有各种各样前所未有的服务出现，非常值得期待。

前文“物联网设备”的说法比较笼统，其实它包括了各种各样的设备。其中具有代表性的便是传感器（图 7.14）。传感器通常是这样一种设备：收集特定的数据信息，并将其转换为电信号供人或设备识别。

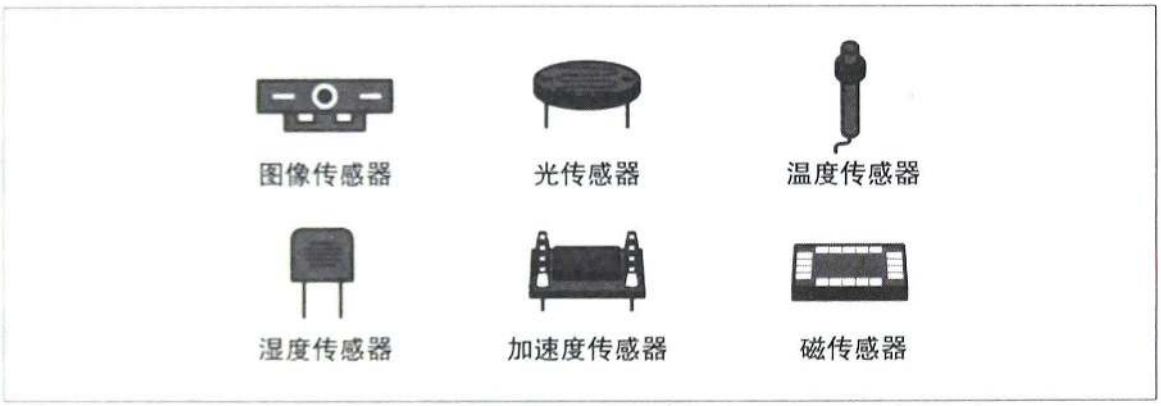
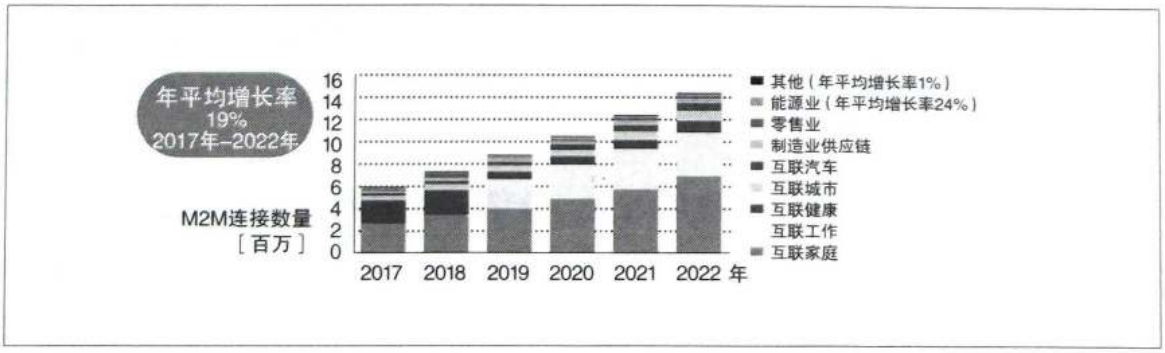


图 7.14 各种各样的传感器

具体来说，传感器中有诸如光传感器（可见光、红外线灯等）、加速度传感器、温度传感器和磁传感器等各种各样的类型，它们能测量的数据各不相同。传感器的构造各有不同，但通常都是将传感器部分与通信模块相连，为其添加通信功能。

物联网的连接数量和通信方式

根据对物联网连接数量变化的预测（图 7.15），物联网设备数量以 20% 左右的年增长率在逐年增加。其中家电、安全等智能手机相关的设备占半数左右，除此以外，互联汽车、智慧城市相关的各领域设备根据预测也会增加。简而言之，物联网的特点便是，大量的设备在各种各样的环境下接入互联网中。



※ 出处：Cisco Systems, Inc., . Cisco Visual Networking Index: Forecast and Trends 2017—2022 [R/OL]. 2018.

图 7.15 物联网连接数量的变化

物联网的通信方式基本上是**无线通信**。其中具有代表性的通信方式是 LoRaWAN、SIGFOX、NB-IoT 等 LPWA 无线通信协议群。这些协议主要分为许可类（NB-IoT 等）和无许可类（LoRaWAN、SIGFOX 等），前者在使用时需要许可，而后者不需要。LPWA 是用于进行低耗电量、长距离数据通信的通信方式，其目标是实现几百米到几千米距离下的通信（图 7.16）。其特点是，为了减少耗电量而控制通信速度，进行几十 Kbit/s 左右的低速间断性通信。LPWA 特意为上行和下行分别设置了发送字节数和发送次数的限制。此外，5G 标准中有多设备连接的相关内容，今后想必也会有物联网设备接入 5G 网络。

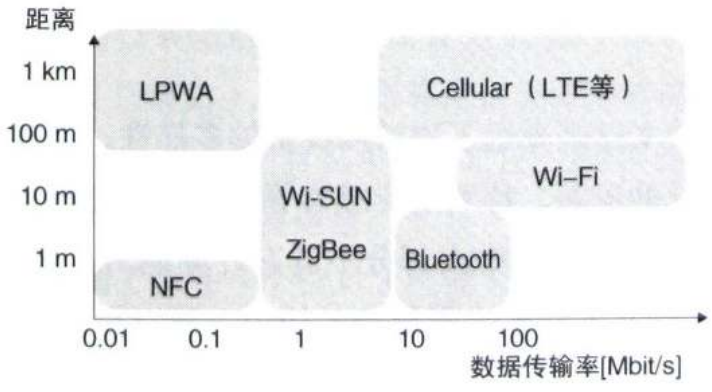


图 7.16 LPWA 的定位

[问题] 处理能力和通信环境上的制约

协议、设备多样性、低速、各种限制……

接下来将从“通信”的视角出发，介绍一下物联网的问题和限制。首先，HTTP 和 MQTT（Message Queuing Telemetry Transport，消息队列遥测传输，图 7.17）是面向物联网的通信协议[1]。

图 7.17 使用 MQTT 时的协议层的构成

HTTP 是在进行 Web 浏览时常用的协议（详见 1.1 节）。MQTT 是工作在 TCP/IP 上层的协议，与 HTTP 等协议相比比较轻量，而且也有可以减少数据量和耗电量的优点。此外，MQTT 不仅适配一对多的通信，也支持异步通信，因此即使客户端没有收到服务器端的应答，也能进行接下来的处理。如上所述，MQTT 拥有多个非常适合物联网设备的特性。

接下来，前文已经介绍了物联网设备的多样性，但其实其中的大多数是处理能力较低的设备。特别是传感器等设备大多比较重视低功耗，而且正因为是传感器，所以其中大部分不具有计算能力。虽然 MQTT 比较轻量，但是由于 TCP 本身是个很复杂的协议，所以要让此类设备使用 TCP 工作，显然处理负担过重。尤其是考虑到安全问题，通常会在 TCP 的基础上使用 TLS（Transport Layer Security，传输层安全协议），但这就对设备的处理能力提出了更严格的要求。

那么，从通信环境的视角来看，又是何种情况呢？在使用 SIGFOX 等

[1] 在主要的物联网平台中，AWS IoT 和 Asure IoT Suite 也支持 HTTP 和 MQTT。

LPWA 算法时，通信有时会特意像第 251 页介绍的那样保持低速，同时还会限制收发数据量和次数。此外，设备如果设置在障碍物较多的房间，则可能出现通信不稳定的情况。

如果在这样的环境下进行通信，就会出现丢包增多、吞吐量下降和时延增大的问题。在这种环境下就可以很容易理解，近些年来开发的适合宽带环境的拥塞控制算法（CUBIC 等）为何完全不合适。

此外，LPWA 的下行流量相比于上行被限制得更多，有时会难收到 ACK，甚至有时要进行 TCP 的 3 次握手都很困难。

此外，还有报告指出，相比于 UDP 首部只有 8 个字节，TCP 首部有 20 个字节，这中间相差的 12 个字节所导致的首部开销过大，以及过去的 *RTO* 计算方法都不适用于物联网。

[TCP 相关动向] 适配物联网 在受限较大的通信中究竟能做些什么

物联网中为活用 TCP 而采取的代表性的手段便是 IETF 的《轻量级实施指南》（Light-Weight Implementaion Guidance，LWIG）之中的《物联网中的 TCP 使用指南》（TCP Usage Guidance in the Internet of Things），该 TCP 技术使用指南目前正在研究和制定中。

举例来说，其中的技术包括将 *MSS* 设置为 1280 字节以下，和推荐使用 ECN 特性等。ECN 是链路上的路由器等设备显式地将拥塞发生这一事件传递出去的功能，其使用 TCP 首部中的 CE 位。ECN 的优点是可以更早地检测到拥塞发生，减少数据发送量。

此外，其中也提到了另外一个针对 3 次握手延迟问题的解决方案，那就是将 TCP 会话维持更长时间。如果无法维持 TCP 会话，那么也推荐使用 RFC 7413 中记述的 TFO（TCP Fast Open，TCP 快速开放）技术。TFO 技术由于同时发送 SYN 和数据，所以会话建立的流程更少。

关于面向物联网的 *RTO* 算法，有意见称 IETF 的 CoRE（Constrained RESTful Environments，受限 RESTful 环境）工作组正在制定的 **CoCoA**（CoAP Congestion Control/Advanced，CoAP 拥塞控制 / 进阶）**机制**比较

合适[1]。

此外，虽然也有人提出了使用 TCP 首部压缩来减少数据量的方法，但由于目前并没有针对物联网的标准算法，所以此方案被搁置，有待后续讨论。

针对物联网设备之间的通信方式受限较大的问题，全世界目前有许多专门的研究和讨论。此外，多种多样的物联网设备和服务，还正处于推广时期，今后想必一定会有新的问题出现，也会有针对这些问题的新的解决方案出现。

针对物联网，也有人在考虑是否可以使用 NIDD（Non-IP Data Delievery，非 IP 数据传输）等 TCP/IP 以外的协议。需要注意的是，无论是哪一种协议，都有其适合或者不适合的环境。换句话说就是各有利弊，因此最重要的是在充分了解各个协议特点的基础之上，根据实际情况来选择合适的协议。

7.4 数据中心

大规模化与各种需求条件并存

数据中心内部网络的高效化是很重要的课题，但其中存在各种各样的需求条件。本节将介绍数据中心内部的网络问题和 TCP 相关的动向。

[背景] 云服务的普及和数据中心的大规模化

从 21 世纪 00 年代后半期开始，云计算提供的服务开使普及，云服务进入了许多普通用户的日常生活之中，其种类简直数不胜数，比如电子邮件、数据存储、群组服务和主机托管等。

在使用这些服务时，用户通过互联网使用由服务商提供的计算资源。

1 Carles Gomez，Andrés Arcia-Moret，Jon Crowcroft. TCP in the Internet of Things: From Ostracism to Prominence [J].IEEE Internet Computing，vol.22，no.1，pp.29-41，2018.

对于用户来说，云计算最大的优势便是使用手边的 PC 和智能手机等设备，就可以很简单地使用各种各样的服务。

为了能够更加高效、稳定地提供此类服务，服务商建造了大规模的数据中心，并在其中集中安装服务器和存储等设备（图 7.18）。

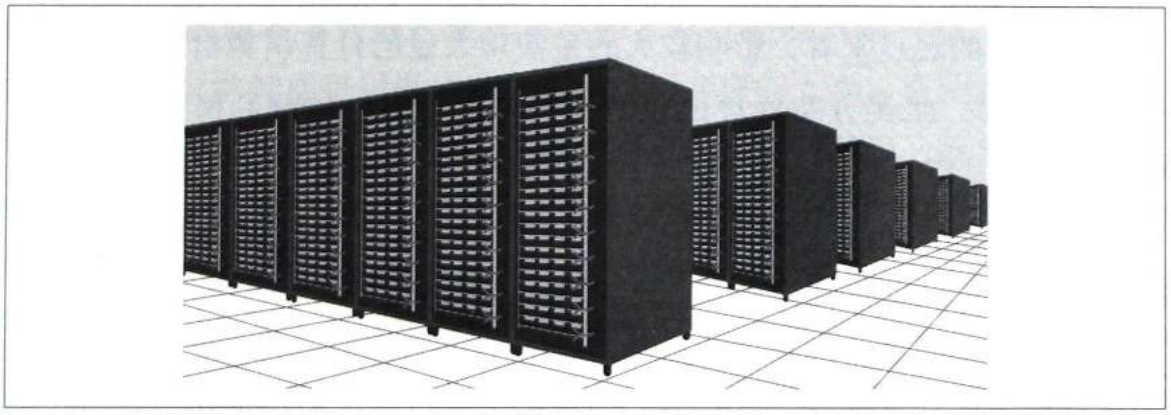
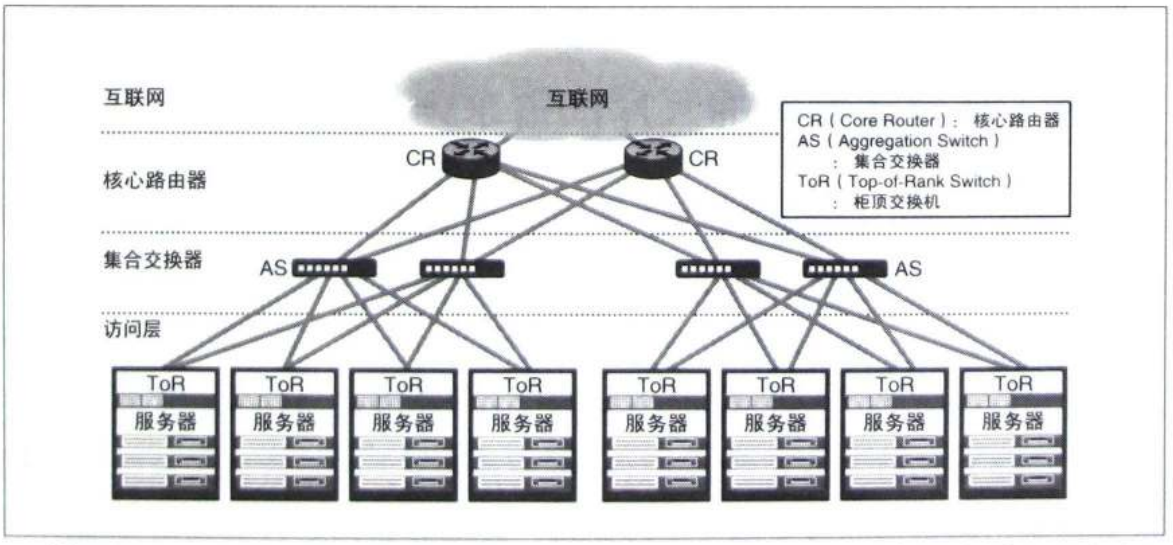


图 7.18 大规模数据中心的示意图

——数据中心内部的网络结构

数据中心内部典型的网络结构如图 7.19 所示。这种网络结构分层次设置了交换机，以便更有效地连接大量的服务器，所以需要大量的交换机设备。



※ 出处：Md. F. Bari, R. Boutaba, R Esteves, et al. Data Center Network Virtualization: A Survey [J]. Communications Surveys & Tutorials, IEEE, vol. 15, no.2, pp. 909–928, 2013.

图 7.19 数据中心内部典型的网络结构

针对此类的网络结构，许多面向数据中心的路由协议被开发出来。这些协议主要以提高负载均衡能力和数据冗余能力为目标。由于本书的主题并非将这些协议网罗来并一一介绍，所以这里主要将重点放在给数据中心内部拥塞发生和拥塞控制带来影响的关键因素，以及与之对应的 TCP 层次的变化情况上。

从上面的观点来看，影响较大的是使用大量的计算机集群进行分布式处理的过程。此类分布处理的过程是，由主机（master）和大量工作机（worker）组成的计算机集群并行处理可以同时处理的任务，以此来实现高速化处理。这些技术中具有代表性的例子就是名为 MapReduce 的算法技术，下面将以此算法为例来进行说明。

此算法的处理流程基本就是重复 Map 逻辑和 Reduce 逻辑。Map 是指主机将输入的任务拆分，然后分配给多台工作机，各台工作机计算分配给自己的任务并得到结果，将其返回给主机。Reduce 则是主机将 Map 过程中收到的数据汇集起来并输出结果数据，然后进入下一阶段。此方法可以通过提高节点个数，进一步提升处理速度，因此在大规模数据处理中十分有效，目前已经变得十分流行。

如果从通信流量的视角来看，MapReduce 算法的特点就是，由主机同时发送数据到各台工作机，然后各台工作机在处理完数据后把相应的结果返回给主机，也就是说，会有**周期性爆发性流量**出现。

[问题] 针对缓冲区的互斥的需求条件

数据中心中一般都有各种各样的服务在运行，也就是说“在数据中心内部，同时有多种多样的流量在服务器之间通信”。而且，不同的服务，其对应流量的特点和进行数据传输时要求的条件也不一样。

其中较为典型的流量类型，主要是“虽然要求低时延，但是数据量比较少”的流量类型，以及“虽然对时延没有要求，但是数据量比较大”的流量类型。如果想同时支持这两种流量，就需要对网络设备和协议提出完全相反的要求。简单来说，如果想确保低时延，就需要减小缓冲区，以防

止各台交换机出现队列时延增大的问题；但如果想准确无误地发送来自数据量较大的网络流的数据包，又需要对各台交换机的缓冲区大小有一定要求。不仅如此，考虑到之前提到的分布式处理时的爆发性流量，也必须确保有一定大小的缓冲区，这样才能提高爆发耐性。

那么，从 TCP 的观点来看，如果要满足这些条件，首先基于丢包的拥塞控制算法就不太合适。这其实与第 6 章介绍的问题一样，其原因就是，基于丢包的拥塞控制算法会一直增大拥塞窗口大小，直到发生数据丢包，这会导致缓冲区被用尽。交换机的缓冲区如果比较小，就会因为缓冲区溢出而频发丢包，而当缓冲区较大时，队列时延就会增大。这样一来，无论如何都无法满足之前提到的两个条件。

此外，即使用上之前提到的 ECN 技术，也只能检测到是否有拥塞发生，无法知道拥塞的程度和持续的时间。如果检测到的拥塞只是瞬时的波动引起的轻度拥塞，就会导致对拥塞窗口大小进行不必要的减小。

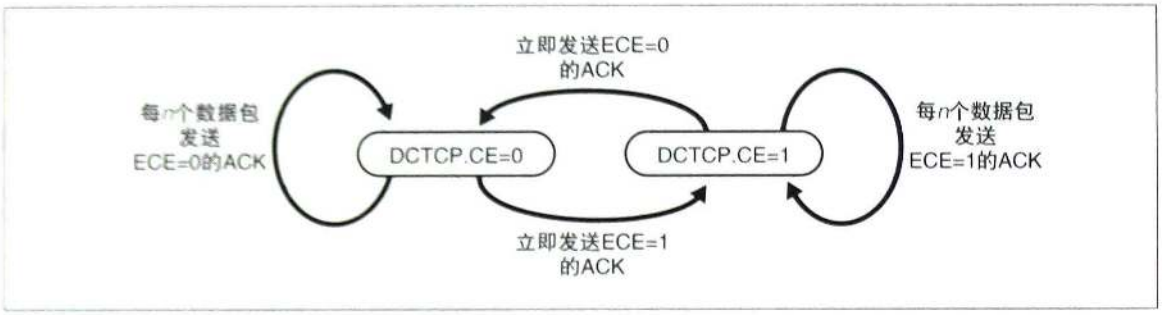
[TCP 相关动向] 面向数据中心的拥塞控制

为了解决上面的问题，DCTCP（Data Center TCP，数据中心 TCP）于 2017 年在 RFC 8257 中完成了标准化。DCTCP 正如其名，是面向数据中心的 TCP 拥塞控制技术，其中制定了 ECN 的扩展技术。简而言之，过去的 ECN 只能简单地检测出是否发生了拥塞，而 DCTCP 会测算发生拥塞的字节比例，然后基于这一比例调节拥塞窗口大小。

具体来说，就是将 DCTCP.CE（DCTCP Congestion Encountered，DCTCP 拥塞遭遇）这一新的比特位变量，用在管理连接状态的 TCB（Transmission Control Block，传输控制块）上。在返回 ACK 时，如果 DCTCP.CE 为 TRUE，就将 TCP 首部中的 ECE（ECN-Echo）标志设为 1。然后根据收到的数据包中代表检测到拥塞的 CE 比特位和 DCTCP.CE 的值，调整返回 ACK 时的行为。当 CE 为 TRUE 且 DCTCP.CE 为 FALSE 时，就将 DCTCP.CE 设为 TRUE，返回 immedidate ACK；当 CE 为 FALSE 且 DCTCP.CE 为 TRUE 时，就将 DCTCP.CE 设置为 FALSE，返回 immediate ACK。

最终的状态变迁如图 7.20 所示。接收 ACK 的发送方使用 *DCTCP.Alpha* 这个新变量来推算发生拥塞的字节比例。使用以下的公式计算 *DCTCP.Alpha*。

$$DCTCP.Alpha = DCTCP.Alpha \times (1-g) + g \times M$$



※ 参考《数据中心 TCP（DCTCP）：面向数据中心的 TCP 拥塞控制》[Data Center TCP (DCTCP): TCP Congestion Control for Data Centers]（RFC 8257）

图 7.20 DCTCP 中 ACK 的生成

在上面的公式中，g 是事前设定好的参数，其取值为 0 和 1 之间的实数。M 代表的是在与 *RTT* 长度一样的观测窗口中接收到的 ACK 之中，ECE 标志被置位的字节占总字节数量的比例。计算出来的 *DCTCP.Alpha* 则用于下面公式中的拥塞窗口大小的计算。换句话说，这个过程就是根据拥塞的程度调整拥塞窗口的大小。

$$cwnd = cwnd \times \left(1 - DCTCP \cdot \frac{Alpha}{2}\right)$$

DCTCP 大致上就是通过以上技术，使用缓冲区大小的小切换来实现高爆发耐性和高吞吐量。但是，DCTCP 显然只可以在数据中心之类的受管理的环境下使用。总而言之，DCTCP 就是为了解决数据中心等特殊环境中发生的问题，基于对其特殊性的考虑而开发出来专有拥塞控制算法。目前 DCTCP 的开发和应用时间仍然很短，今后想必一定会有更加广泛的应用和改良。

7.5 自动驾驶

追求高可靠性与低时延、大容量的通信性能

以规避冲突等辅助驾驶功能为契机，汽车的发展进一步加速。为了支持自动驾驶，不仅车辆自身的性能比较重要，通信也发挥着相当重要的作用。下面将简单介绍实现自动驾驶的技术，以及其与 TCP 之间的关系。

[背景] 以普及自动驾驶为目的的技术

自动驾驶指的是由系统来完成人类驾驶时进行的各种行为（**认知、判断、操作**）。其核心技术便是通过 GNSS（Global Navigation Satellite System，全球导航卫星系统）、摄像头、雷达和传感器等传感设备和信息通信技术，一边识别道路形态、移动物体（车辆、行人），以及建筑物等周边环境，一边进行自动驾驶控制。

——自动驾驶分级

自动驾驶按照难易度分为若干个等级。表 7.1 展示的便是由美国非营利组织 SAE（Society of Automotive Engineers，美国汽车工程师协会）定义的自动驾驶分级的例子。

表中定义了等级 0 到等级 5 总共 6 个级别：在等级 0 到等级 2，驾驶者是主体，部分驾驶任务交由车辆系统辅助完成；而在等级 3 到等级 5，由系统作为主体完成驾驶任务，虽然当系统很难完成任务或者遇到危险的情况时，需要驾驶者介入进行处理，但其实等级 5 的目标是完全自动驾驶。从等级 3 开始，系统需要感知和预测车辆的行驶情况和危险，并将结果数据反馈到驾驶上，因此技术上的实现难度很高。

表 7.1 自动驾驶分级

等级	概要	与安全驾驶相关的监控、处理主体
驾驶人完成部分或者全部动态驾驶任务		
等级 0	· 驾驶人完成全部动态驾驶任务	驾驶人 无自动驾驶
等级 1	· 系统仅在限定范围内完成横向或者纵向车辆运行控制子任务	驾驶人 辅助驾驶
等级 2	· 系统在限定范围内完成横向和纵向车辆运行控制子任务	驾驶人 部分自动驾驶
自动驾驶系统（工作时）完成全部动态驾驶任务		
等级 3	· 系统在限定范围内完成全部动态驾驶任务 · 在难以继续工作时，驾驶人需要响应系统的介入请求等	系统 有条件自动驾驶 （在无法工作时由驾驶人驾驶）
等级 4	系统完成全部动态驾驶任务，并在限定范围内完成在难以继续工作时的处理任务	系统 高度自动驾驶
等级 5	系统完成全部的动态驾驶任务，同时不受限地（即在非限定范围内）完成难以继续工作时的处理任务	系统 完全自动驾驶

※ 动态驾驶任务（DynamicDriving Task，DDT，J3016 相关用语的定义）：
在道路交通中，除行程规划和目的地选择等策略上的功能以外，驾驶车辆之时需要实时完成的所有操作和决策上的功能，包含但不限于以下子任务。
1. 控制方向盘进行横向的车辆运动控制。
2. 通过加速或减速进行纵向的车辆运动控制。

※ 出处：《日本官民 ITS 构想路线图 2018》

不同的国家、不同的机构在详细的等级划分方面并不完全一致，因此需要进行统一整理，但是在驾驶人究竟在多大程度上参与到驾驶中这个方面，以及事故发生时责任如何划分这个方面，目前业内正在进行详细且严肃的讨论。近些年来，冲突回避和加速或方向盘控制等部分自动系统作为初级阶段的辅助驾驶手段开始被引入到当前上市的车辆中。目前，政府和民众正在齐心协力推进完全自动驾驶的实现，其中也包括法律方面。

无线通信所承担的职能 V2N、V2V、V2I、V2X、V2P

要实现自动驾驶，无线通信毫无疑问承担着十分重要的作用。近年已有使用移动网络进行通信（Vehicle-to-cellular Network，V2N），并进行软件升级等的汽车上市销售。几年以后，车辆间通信（Vehicle-To-Vehicle，

V2V）和车辆道路间通信（Vehicle-to-Infrastructure，V2I）等形态的通信想必也一定会普及。要想规避与行人之间的冲突，人车之间的通信（Vehicle-to-Pedestrain，V2P）也十分重要。这些技术总称为 V2X（Vehicle-to-Everything）（图 7.21）。

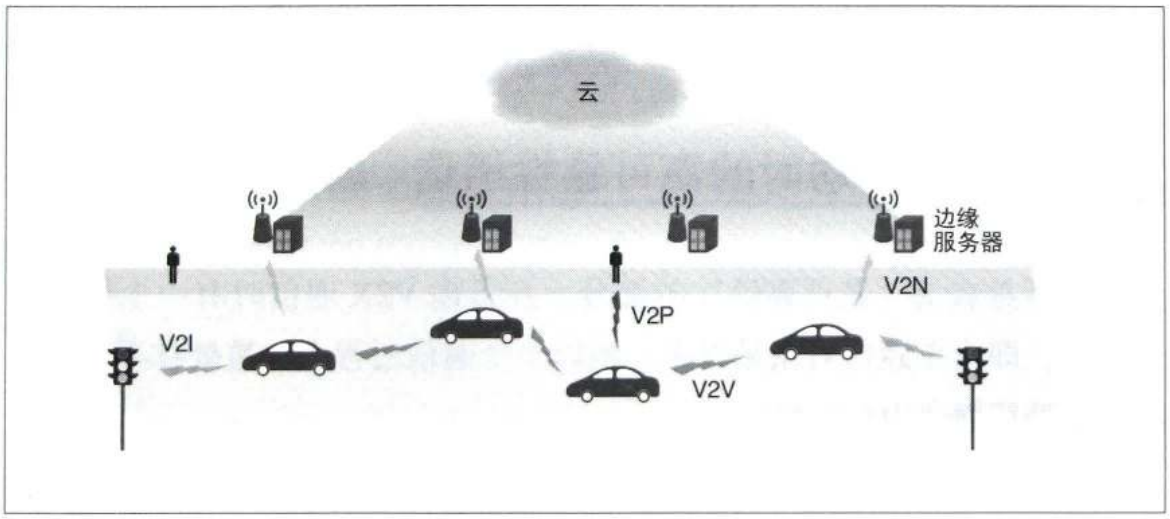


图 7.21 自动驾驶通信

对通信性能的要求 高可靠性、低时延、大容量

近些年来，摄像头和传感器已经成为汽车的标配，这些设备主要用来收集自己所能认知范围内的数据信息。如果将这些信息与利用 V2X 技术通信所得到的数据合并起来进行分析，一定可以得到远超原先范围、更加广阔的四周交通情况。按照这样的趋势继续发展下去，就可以期待出现更高程度的自动驾驶。自动驾驶汽车要想做到及时获取与周边情况相关的数据并将数据反馈到驾驶控制中，显然就必须要有高可靠性、低时延、大容量的通信性能，这些正是 5G 中提出的通信性能。

在“大容量”方面，如果考虑到迄今为止的无线通信技术的发展情况，那么可以认为其实现的可能性相当高。难题在于如何实现“低时延”。

目前 MEC 技术被认为十分有前景。MEC 是指，将服务器配置在本地区域内，并让流量只在本地区域内流动，与此同时通过本地服务器完成一系列的处理运算。

在通常情况下，通信链路是移动设备→互联网→云服务器→互联网→移动设备，但在引入 MEC 技术之后，通信链路可以简化为移动设备→边缘服务器→移动设备，因此这项技术可以说实现了端到端的低时延化（图 7.21）。MEC 不仅有助于减少通信时延，而且由于其减少了互联网上不必要的数据发送，所以也可以减少整个互联网上的网络流量。

[问题] 高速移动时的高可靠性通信

把握实现自动驾驶的命脉——可靠性

最后是有关“高可靠性”的部分。在考虑 V2X 通信时有一个情况需要注意，即由于通信对象是汽车，所以整个通信过程会**一直处在不稳定的高速移动环境**之中。

在移动的环境中，电波不仅会发生激烈变化，而且还会与声波一样产生多普勒效应。毫无疑问，如何使这种严苛环境中的通信更加稳定，便是使用 V2X 实现自动驾驶的最核心的问题。

补偿劣化信号的信号处理技术，以及选择与哪辆汽车或者基站进行通信等多个领域相关的研究，目前都在如火如荼地进行中。

3GPP Release 14 中制定了针对 V2X 的技术文档。其中，关于时延方面的主要要求，V2P/V2P/V2I 是 100 ms，而 V2N 是 1 秒。虽然与 5G 的目标 1 ms 相比，这些时延要求较大，但由于其可以利用现有的 4G 基础设施，以及无线区域目前已经普及且能覆盖更为广阔的区域，所以也一定会有许多适合的应用场景。

使用 5G 网络进行研究的示例之一，便是进行如下的实证实验：使用 V2X 来完成卡车的队列行进过程的自动化（图 7.22）。经由互联网使用 V2N 技术，进行远距离的监视和控制，实现卡车之间的车辆间通信。通常来说，使用某个通信运营商的网络这类闭环网络进行控制是比较现实的。然而如果要进行远程控制，可能必须要经过外部网络。

如果使用外部网络，就必须考虑拥塞。此时，“低时延和可靠性两者兼得”显然是一个非常重大的课题。