

03-视频 互联网介绍

互联网是如何工作的？

了解互联网的基础知识以及此短视频系列相关的所有内容

基于Web应用程序的爆炸式增长，每个开发人员都可以从理解Internet的工作原理中受益。在本文中，通过code.org提供的一系列介绍Internet的短视频，你可以了解到Internet的基础知识以及它是如何工作的。通过这篇文章，每个人都能回答以下问题：

- 互联网是什么？
- 在互联网中信息是如何传输的？
- 网络如何互相通信以及所设计的协议？
- 数据包、路由器和可靠性之间的关系是什么？
- HTTP和HTML--你如何在浏览器中查看网页？
- 互联网上的信息传输是如何确保安全的？
- 什么是网络完全，哪些是常见的互联网犯罪？

互联网是什么？

互联网是一个由计算机相互连接的全球网络，它们通过一组标准化的协议进行通信。

在下面的视频中，“互联网之父”之一Vint Cerf解释了互联网如何运作的历史，以及为何没有一个人或组织能真正负责它。

Vint Cerf: 很多很多年以前，在1970初期，我的伙伴，鲍勃卡恩(Bob Kahn)和我开始着手于现在被我们称作互联网的设计，这是另一项实验的成果叫ARPANT(Advanced Research Projects Agency NETwork)，意思是：高等研究计划部署网络，它曾是国防部（美国）的研究计划。巴兰(Paul Baran)尝试想建立一个能承受核弹攻击的通讯系统。所以他的点子是，把信息拆分成一个个小区块，然后尽快地利用各种可能的方向，透过网状网络（Mesh Network）来发送。所以我们建出了这个，最后变成了全国性的实验封包网络，并且成功了。

有谁控制国际网络吗？

诚实的答案是，嗯...没有人控制国际网络。或者说，每个人都可以控制国际网络，而且独立运作的网络所组成。国际网络最有趣的，就是它是一个天然的分布式系统，并没有集中化的控制来决定封包的传递路径，或是个别网络构造的位置，或是谁能跟谁连接。这些都是由网络作业员各自独立决定，网络作业员都意图去确保网络里每个部分点对点（end-to-end）的连接性。因为国际网络的使用效益靠的就是任何装置都可以跟其它装置沟通，就像你想要能跟世界上的任何人通电话一样，这是前所未有的做法。这个理念：你知道的东西或许可能对别人极为有用，反之亦然（别人知道的东西对你可能很有用）。这个理念构成非常强大的驱动力，让人们分享信息的驱动力。顺带一提，这正是科学发展所需要的基石：人们分享信息。

有了国际网络，我们开始思考创造新的应用，也许是在手机上编写一个app应用程序，也许是基础设施，让还没有网络的人也能使用网络，或许只是单纯的每天上网。既然你无法避免不接触网络，那么为何不去了解它并且利用它呢？

电线、电缆和Wi-Fi

Internet 上的信息以比特的形式通过各种介质从计算机传输到另一台计算机，包括以太网电缆(电话线)、光纤电缆(网线)和无线信号(Wifi)（即无线电波）。

在下面的视频中，你将了解互联网上不同的数据传输媒介，以及每种媒介的优缺点。

我是谷歌的软件工程师Tess Winlock。

那么问题来了：图片，短信或者电子邮件是怎么从一台设备传递到另一台设备的呢？不是靠魔法，而是通过互联网，一个用来传递信息的现实物理系统。

互联网很类似我们的邮政服务，只不过在传送的物理层面上有点儿不同。互联网不使用邮箱和信封，而是用二进制数据来传递信息，所有数据都由位（电脑信息单位）组成。一个（bit）位可以用来表示任何一组相对意义的语词，例如：开和关、是和否。数位信息世界通常使用1代表开，0表示关。因为一个位元码可以表示2种状态，所以称为二进制码。8个位元合在一起组成一个字节（Byte）。1000个字节称作千位元组（缩写为KB）1000个千位元组称作百万字节（缩写为MB）而常来说一首数位音乐大约使用3~4百万字节（3MB~4MB的数据所组成。不论传送的是图片、视频或是一首歌，所有在互联网中传送的数据都简化成位码（bits）。位码（bits）也是数位信息的最小单位。但和我们现实生活中要从甲地发送1和0信息到乙地不同，所以，在互联网透过有线或无线方式传送信息时，到底发生了甚么呢？

我们先用一个实际例子来说明，在生活中我们要如何从甲地发送一个位（bit）讯到乙地 我们用开灯表示1，关灯表示0，这个方式来说明。当然也可以利用声音或其他方式代替像是摩斯电码。只是这些透过人工传递的方法，传递信息又慢又容易出错。而且非常倚靠人力，我们真正需要的是透过机器传递。经过不断演进，人类已经制造出许多系统可以透过实体媒介传递不同形式的二进制数据。

今天我们已经利用电力、光和无线电波传递二进制数据，想像一下现在有两个灯泡透过一条铜线连接着，我们就可以传送二进制数据了。使用开关控制电力让两边的灯泡发亮关闭电力则灯泡熄灭（没有电力则没有灯亮）我们假设两边以灯泡发亮表示1，灯泡熄灭表示0，这样我们就有一个透过电力发送二进制数据的系统了。不过，我们遇到了一个小问题，万一你刚好要传送五个0，要如何让两边的人可以正确地传送和接收到正确的数量呢？有个解决办法是利用定时器。灯泡两端的人可以说好每一秒钟 传送一个位的数据，接收端就可以每一秒钟记录一次。所以，要传送五个0，就是关灯五秒钟 这样接收者就可以记录到五秒钟的位数据。如果要传送五个1，就是打开电灯持续五秒钟，接收者就按照时间纪录。

可是，这样的传送速度有点慢，因此我们需要增加传送频宽 频宽（传输媒介的最大传输数据量）的单位是位速率（Bitrate），指的是在单位时间内可以传送的二进制数据量，通常以秒为单位。

另一个测量速度的单位是延迟时间（latency），或是传送一位所花的时间 延迟时间是一个位从来源端送出后，到目的端接收到这个位，中间所花的时间。对人类而言，一秒钟传送一个位数据是很快速的 比我们用开关灯的方式还快多了！实际来说就是下载3MB的歌曲只花三秒钟，或是8Mbps就是说每一秒钟可以传送8百万个位元资料 显然地，透过人力传递是无法做到8Mbps的，但是透过机器传递，就可以妥善完成。不过我们会遇到其他的问题，就是要用哪种线或者互联网可以传多远。如果透过网路线传输 就是家里、学校或办公室常见到的那种网路线，大概一百公尺远就会遇到信号遗失或干扰的问题。

可是互联网必须要全世界使用，我们需要一个更好的办法来传送更远的距离。甚至传到海洋的另一端。我们到底该用甚么呢？还有甚么东西可以传递的比电力还要快呢？光，就是它了！我们可以利用光来传递信息 透过光纤电缆（fiber optic cable）。这是一种利用玻璃制作来传递光线的材料，当我们使用光纤来送光束时，光信号在光纤中不断前进，直到光信号被接收端接收。根据传送角度不同 还可以同时传送多组信号，所有信号都以光速前进。光纤的速度非常非常快，更重要的是信号衰减很低 所以我们可以传递数位信号几百英里也没有信号遗失 因此我们把光纤电缆安装在海底，连接海洋彼岸的两地。在2008年时，埃及亚历山大港附近的海底光纤电缆断了，导致中东和埃及互联网中断。虽然我们把互联网视为理所当然，但其实它是很脆弱的系统。而且，光纤非常棒，但也十分昂贵和费工。所以许多时候你会用铜缆来替代。

那我们又是如何不用线路传递信号呢？无线传递信息又是怎么办到的？无线传讯一般来说是用无线电波信号来传递 这些设备必须把一堆的1和0信号转换成无线电波的不同频率 接收端的设备接收到电波之后，再把无线电波转换成电脑可以处理的二进制数据透过无线技术让网络可以方便移动。但无线电波信号没有办法准确地传递正确信号到足够远的地方。就像是你没有办法在洛杉矶收听到芝加哥的电台节目一样。虽然无线传输很方便但它还是依靠有线网络。假设你在咖啡店使用无线网络，你所传送的二进制数据 会被传到无线路由器，再透过网路线等实体线路在互联网世界中传递 到世界各地。透过网路线传递信息的方法可能会日益进步 不论是利用卫星传递激光光束、无线电波、热气球或是用无人机。最基础的二进制数据传递协定，用来表示各种信息或是传递数据的通信协定 还是没多大的改变。互联网世界的所有数据，不管是文字、电子邮件、图片或视频，所有的信息都会被转换 变成无数的1和0，再用电子脉冲、光束、无线电波，传递所有信息。

IP Adresses(IP地址) and DNS(域名)

现在我们已经了解了通过Internet(互联网)传输数据的物理介质（第二天的笔记），是时候了解所涉及的协议了。在这个庞大的全球计算机网络中，信息如何一台计算机传输到另一台计算机？

下面将简要介绍IP、DNS以及这些协议如何使Internet工作的！

互联网协议地址(IP Adress)(Internet Protocol Adress)、域名名称系统(DNS)(Domain Name System)。

回溯到1970年代的时候，还没有标准的做法，在Vint Cerf与Bob Khan发明了网络通信协议之后，才让互联网这样的沟通方式得以实现，这项发明为我们所说的成为了互联网的重要基础，它将地球上数十亿的设备连接到一起。

你可能是使用笔记本电脑或手机透过Wi-Fi上网，Wi-Fi则连接到互联网服务提供商（ISP），ISP则透过数十万互相连接的网络让你与世界上数十亿的装置得以连接。

多数人没有注意到的一点是：互联网其实是一种设计哲学，互联网也是由一组协议来表现的结构体系 一个通信协议是由一组大家公认的规则 and 标准组成 如果所有参与者都同意使用，就能让彼此顺利的沟通。互联网「实际上」如何运作，反而没那么重要。更重要的是：它的设计哲学能够让互联网调整与纳入新的通信技术，这是因为新的通信技术只需要配合协议，就能以某种方式使用互联网。

各式各样的设备在互联网世界里，都拥有一个独一无二的地址，这些网络地址就是一组数字，很类似电话号码或街道地址。这个网络地址，在一个范围的网络内，对每个电脑或设备而言都是独一无二的。

这就好比大多数的家庭或公司都有邮件地址。你不需要认识对方，就能寄信给他，但你需要知道他的地址，以及该如何正确的书写地址，如此一来，这封信才能透过邮政系统送达目的地。在互联网里，电脑的地址系统有类似的作用，这形成了最重要的通信协议之一：互联网协议（Internet Protocol），简称：IP。电脑的地址也因此称为：IP地址。

浏览网站等于是：你的电脑向另一部电脑发送信息。你的电脑向另一台电脑的IP地址发送信息，同时也附带着自己的IP地址。所以，另一台电脑就知道要回复信息到什么地方。

你也许有看过IP地址，你它其实就是一连串的数字！这些数字是由「层级」所构成，就像家庭地址有：国家、城市、街道、门牌号码那样。IP地址也可分为许多部分，与所有的数位数据一样，每一个数字都是以位（bit）所表示，典型的IP地址的长度是32位，每一个部分则有8位。

前面的数字，通常可识别装置所在的国家或区域，接下来表示：子网域，最后则表示：特定装置的地址。这个版本的IP协议，称为IPv4，

它是在1973年所设计，并且在80年大初期被广泛使用。提供了超过40亿个独一无二的地址，让装置连接到互联网。

但是互联网远比Vint Cerf想像的更为普遍40亿个不同的地址仍然不够！我们正处于转换到更长的IP协议（IPv6）的过渡期中，IPv6的每一个地址的长度为128位，可提供超过 340×10^{33} 个不同的地址，这个数量多到足够让地球上每一颗沙子都有自己的IP地址！

大多数的人不曾见过或关心IP网络地址，域名名称系统（DNS）可让网址名称（www.example.com）与对应的IP地址关联在一起，你的电脑会使用DNS查询网域名称，来取得对应的IP地址，IP地址则用来让你的电脑，能够连接到互联网上的目的地。

看个例子... (声音一：嗨！我想要前往www.code.org) (服务器一：嗯...我不知道那个网域的IP地址，让我问一下) (服务器一：嗨，有人知道code.org要怎么去吗？) (服务器二：耶，我知道，它就是在174.129.14.120) (服务器一：喔！赞，谢拉！我要把它记下来保存，以免之后我又需要它) (服务器一：嘿！这是你需要的IP地址) (声音一：好喔！谢谢拉！)

要如何设计一个系统能让：数十亿个装置，都能在茫茫网站大海中，找到正确的那一个呢？只有一台DNS服务器，一定无法处理所有装置的查询请求。解决方法是：让所有的DNS服务器以「分布式阶层」串联在一起，并且以区域分开，针对主要的网域 (.org, .com, .net) 分开负责，DNS起初是为了政府与教育机关而设计的开放且公开的通信协议，因为DNS的开放特性，它很容易受到网络攻击的影响。有一种攻击称为：伪造DNS纪录 (DNS spoofing) 这是指：骇客入侵一台DNS服务器，并且将网域名称修改对应到错误的IP地址如此，攻击者就能把人们传送到「冒名顶替」的假网站 如果这发生在你身上，你会信以为真的使用这个假网站，让你受到更多问题的伤害。互联网非常巨大，且每天都不断扩大。然而，DNS以及互联网协议都被设计可以持续扩充。无论互联网会成长到多么巨大！

数据包、路由和可靠性

Internet 上的信息从一台计算机传输到另一台计算机不需要遵循固定路径；事实上，它可能会在传输过程中改变路径。此信息传输以数据包的形式完成，这些数据包可能会根据某些因素遵循不同的路线。

在本视频中，您将了解信息包如何从一台计算机路由到另一台计算机以到达目的地。

嗨，我的名字是林恩。我是Spotify的软件工程师。我会第一个承认，我经常认为互联网的可靠性是理所当然的。然而，互联网的信息量巨大，要如何可靠地将每个数据传递给你呢？假设您想播放Spotify的歌曲，你的计算机似乎是直接连接到Spotify服务器，Spotify会在专用线路上直接向你发送歌曲。但实际上这不是互联网的运作方式。如果互联网是由直接的专用连接构成，那么随着数百万用户的加入，就不可能保持工作。特别是不能保证每根电线和电脑都能正常工作。相反地，数据是以更不直接的方式在互联网上传播。

很多年前，在20世纪70年代早期，我的合作伙伴Bob Kahn和我开始着手设计我们现在称之为互联网的东西。Bob和我有责任和设计互联网协议及其架构。因此，我们始终坚持参与互联网的增长和发展，直到现在也在继续。

信息从一台计算机传输到另一台计算机的方式非常有趣。它不需要遵循固定的路径，事实上，你的路径可能会在计算机中转换为计算机对话。互联网上的信息在我们所说的信息包中，从一台计算机传到另一台计算机，一个信息包在互联网上从一个地方传播到另一个地方，就像你在车里从一个地方到另一个地方一样。根据交通拥堵或道路状况，每次旅行时，您可能会选择或被迫采取不同的路线到达同一个地方。就像你可以在一辆车里传送各种各样的东西一样，很多数字信息可以用IP包发送，但也有一些限制。如果你需要将航天飞机从其建造地点移动到发射的位置，该怎么办？因为航天飞机不能装在一辆卡车里，所以需要将它分解成碎片，用一队卡车运输。他们都可以采取不同的路线，并可以在不同的时间到达目的地。但一旦所有的碎片都到达后，你就可以重新组装成完整的航天飞机并准备发射。

在互联网上，细节同样起作用。如果你要将非常大的图像发送给朋友或上传到网站，图像可能由数以百万计的1和0比特组成，数量太大，无法在一个包中发送。由于它是计算机上的数据，图像可快速被分解成数百甚至数千个称为数据包的较小的部分。与汽车或卡车不同，这些数据包没有司机，也没有选择路线。每个包都有它的来源和去向的互联网地址。互联网上，称为路由器的特殊计算机就像流量管理器一样，可以使数据包顺利通过网络。如果一条线路拥堵，个别的数据包可能通过互联网通过不同的路线，它们可能在稍微不同的时间或者不同的顺序到达目的地。

我们来谈谈“路由器”是如何工作的。作为互联网协议的一部分，每个路由器都跟踪用于发送数据包的多条路径，并且它根据数据包的目的IP地址为每个数据选择最便宜的可用路径。在这种情况下最便宜并不意味着成本，而是时间和政治等非技术因素和公司之间的关系。通常，数据传输的最佳路径并不一定是最直接的。具有路径的选项会使网络容错。这意味着这意味着即使出现可怕的错误，网络也可以继续发送数据包。这是互联网关键原则的基础：可靠性。

现在，如果你想请求一些数据而不是所有的数据都交付了呢？假如你想听一首歌。你如何100%确定所有数据都将被传送，以便歌曲完美播放？这就需要介绍你最好的朋友TCP（传输控制协议）了。TCP以数据包的形式管理所有数据的发送和接收，可以把它想象成一个有保障的邮件服务。当你在设备上请求歌曲时，Spotify会将一首歌分解为多个数据包。当数据包到达时，TCP会进行完整清点并回复以确认收到

每个数据包。如果所有数据包都在那里，TCP会签收您的交付，您就成功收到了歌曲。如果TCP发现某些数据包丢失，它将不会签收否则你的歌听起来不会悦耳，或者歌曲的某些部分可能会丢失。对于每个丢失或不完整的数据包，Spotify将重新发送它们。一旦TCP验证了该歌曲请求的多个数据包的传递，您的歌曲就会开始播放。TCP和路由器系统的优点在于它们是可扩展的。他们可以使用8到80亿台设备。实际上，由于这些容错和冗余原则，我们添加的路由器越多，互联网就越可靠。

更棒的是我们还可以在不中断它服务的情况下增长和扩展互联网。互联网由数十万个网络和数十亿台物理连接的计算机和设备组成。这些组成互联网的不同系统相互连接，相互通信，并协同工作，这是因为在互联网上传输数据的方式是一致的。计算机设备，或者互联网上的路由器，帮助所有的数据包到达目的地，如果有必要的话，在那里重新组装。这种情况每天发生数十亿次，你和其他人是否正在发送电子邮件，访问网页，进行视频聊天，使用移动应用程序，或互联网上的传感器或设备相互通信时。

HTTP和HTML

HTTP是网页在互联网上传输的标准协议。下面的视频是一个简单的介绍HTTP和web浏览器如何加载网站。

互联网如何运作：HTTP和HTML

我是Jasmine，我是XBOX One工程团队的项目经理。我们最大的功能之一是XBOX Live。这是一项依靠互联网连接世界各地玩家的在线服务。这不是一件容易的事，幕后发生了很多事情。

互联网完全改变了人们互动和联系的方式。但它是如何工作的？世界各地的计算机又是如何相互交流的？让我们来看看网页浏览。

首先，打开Web浏览器，这是你用来访问网页的应用程序。接下来，你键入Web地址或URL(Uniform Resource Locator资源统一标识符)，它代表你想访问的网站的统一资源定位器，如tumblr.com。

嗨，我是大卫卡普，Tumblr的创始人，今天在这里谈谈我们每天使用的网络浏览器实际上是如何工作的。你可能想知道当你在浏览器中输入一个地址www.example.com，并点击Enter之后，实际发生了什么呢？这真是你能想象到的最疯狂的事情。从那一刻起，你的计算机就开始与称为服务器的另一台计算机通话，这个距离通常是数千英里。并且在几毫秒内，你的计算机会向该服务器请求网站服务器以称为HTTP的语言开始与您的计算机对话。HTTP代表超文本传输协议，你可以把它看作是一台计算机用来向另一台计算机索取文档的语言。而且它实际上非常简单。如果你要截取计算机与互联网上的Web服务器之间的对话，它主要是由称为“GET”请求的东西组成。因此，如果你尝试登录Tumblr并加载我们的登录页面，你所要做的就是向Tumblr的服务器发送一个GET请求，上面写着get/login。这告诉Tumblr的服务器，你需要Tumblr登录页面的所有HTML代码。所以HTML(HYPER TEXT MARKUP LANGUAGE)代表超文本标记语言，你可以将其视为用于告诉Web浏览器如何使页面外观的语言。如果你想到像维基百科这样的东西，它实际上只是一个简单的大文档，HTML是你用来使标题变大和粗体的语言，使字体成为正确的字体，将某些文本链接到其他页面，使某些文本粗体化，使一些文本斜体，将图像放在页面的中间，使图像与右边对齐，使图像与左边对齐。

网页文本是直接包含在HTML中的，但是图像或视频等其他部分是具有自己需要的URL单独档案的。浏览器向其中的每一个发送单独的HTTP请求，并在它们到达时显示它们。如果一个网页有很多不同的图像，并且每个图像都会产生一个单独的HTTP请求这会使页面加载速度减慢。现在，有时候当你浏览网页时，你不仅仅是在请求带有GET请求的页面。有时你发送信息就像填写表格或输入搜索查询一样，浏览器使用HTTP POST请求以纯文本形式将此信息发送到Web服务器。

假设你登录Tumblr，那么第一件事是你提出登录请求，也就是tumblr的登录页面，并且有一些附加数据。它将有你的电子邮件地址和密码。那会把您带领到Tumblr的服务器。Tumblr的服务器知道了你就是大卫。于是，它把一个网页发回给你的浏览器，上面写着“成功”，登录者为大卫！但是除了网页之外，它还附加了一些不可见的cookie数据，你的浏览器可以看到并知道如何保存这些数据。这非常重要，因为它确实是网站记住你是谁的唯一方式。并且你的Web浏览器会保留该号码，并在你下次刷新Tumblr时，或去Tumblr.com时，你的网络浏览器会自动知道将该ID号附加到它发送给Tumblr服务器的请求中。所以Tumblr的服务器可以看到来自你的浏览器的请求及ID号，并且知道，“好的，这是大卫的请求。”

现在，互联网是完全开放的。它的所有连接都是共享的，信息以纯文本形式发送。这使得黑客能够窥探

到你通过互联网发送的任何个人信息。但是安全网站阻止了这一点，通过要求你的Web浏览器使用安全套接字层及其后续传输层安全性在安全通道上进行通信。你可以将SSL和TLS视为围绕通信的安全层保护他们免受窥探或篡改。当你看到浏览器地址栏中显示在HTTPS旁边的小锁时，SSL和TLS是活动的。HTTPS协议确保你的HTTP请求是安全的、受到保护的。当网站询问您的浏览器时，为了进行安全连接，网站首先提供数字证书。这就像一张官方身份证，证明它是它声称的网站。数字证书由证书颁发机构发布，证书颁发机构是验证网站身份并为其颁发证书的受信任实体，就像政府可以签发身份证或护照一样。现在，如果网站尝试在没有正确发布的数字证书的情况下启动安全连接，你的浏览器会发出警告。这是网页浏览的基础知识！也是我们日常看到的情况。

总而言之，HTTP和DNS管理HTML、媒体文件或Web上的任何内容的发送和接收。让这成为可能的是，TCP/IP和路由器网络，它们以小数据包的形式分解和传输信息。这些数据包本身由二进制序列组成，序列分别为1和0，并通过电线，光纤电缆和无线网络实际发送。一旦你学会了互联网的工作原理，你就可以依赖它，而不用记住所有的细节。我们可以相信，所有这些层将一起工作，以规模和可靠性先后交付信息。