

白皮书

实时网络视频监控中的延迟

六月 2024

概述

互联网协议 (IP) 网络视频监控中的流媒体直播过程包括在摄像机中捕捉视频、通过网络打包传输以及在接收器中解包显示。每个步骤都会或多或少地增加延迟。

- **摄像机端引起的延迟。** 捕捉视频之后要进行图像处理、压缩和打包。每个步骤都会带来一定的延迟，但总体而言，摄像机内的流程只占端到端总延迟的一小部分。
- **网络引起的延迟。** 它可能非常大，也可能非常小，这是端到端延迟“等式”中非常不可预测的因素。您可以通过投资良好的网络软件和硬件，使其更加可预测。网络延迟在很大程度上取决于数据与带宽的比率。您可以对摄像机进行配置，减少其生成的数据量，从而减少需要通过网络发送的数据包数量。
- **客户端引起的延迟。** 在客户端，数据被接收并缓冲排序，然后排队输出到显卡和监视器。客户端的接收缓冲区对延迟的影响更大，甚至可长达数秒。如果没有大缓冲区，视频流就有可能无法均匀播放。

减少延迟始终是一个成本问题。改进网络和客户端的硬件与软件可以取得更大的成功。

目录

1	引言	4
2	什么是延迟?	4
3	如何测量延迟?	4
4	什么会影响延迟?	5
	4.1 摄像机引起的延迟	5
	4.2 网络延迟	7
	4.3 客户端延迟	9
5	减少延迟	10
	5.1 减少摄像机延迟	10
	5.2 减少网络延迟	11
	5.3 减少客户端延迟	11

1 引言

在视频监控环境中，延迟是指从捕捉帧的瞬间到显示同一帧的瞬间之间经历的时间。这也称为端到端延迟或传感器到屏幕延迟。从摄像机传感器向显示监视器传输帧的过程涉及很长的一系列步骤。

本白皮书概述造成总延迟的各种步骤。它还就如何减少延迟提出建议。

2 什么是延迟？

延迟的定义取决于上下文。在网络技术中，延迟通常被认为是从源端发送一条信息到最终目的地接收到同一条信息之间的时间延迟。

本文将讨论网络视频监控系统中的延迟问题。在这里，我们将延迟定义为从摄像机捕捉图像到图像在视频显示屏上可见之间的时间延迟。在这段时间内，图像被采集、压缩、传输、解压缩和显示。每个步骤都会在总延迟中增加自己的延迟份额。为简单起见，端到端延迟可视为由三个主要阶段组成：

- 摄像机引起的延迟（图像处理延迟、压缩延迟）
- 网络引起的延迟（传输延迟）
- 接收端引起的延迟（客户端缓冲、解压缩延迟、显示延迟）

为了达到视频监控系统的延迟目标，在设计视频解决方案时必须考虑上述每种延迟。

3 如何测量延迟？

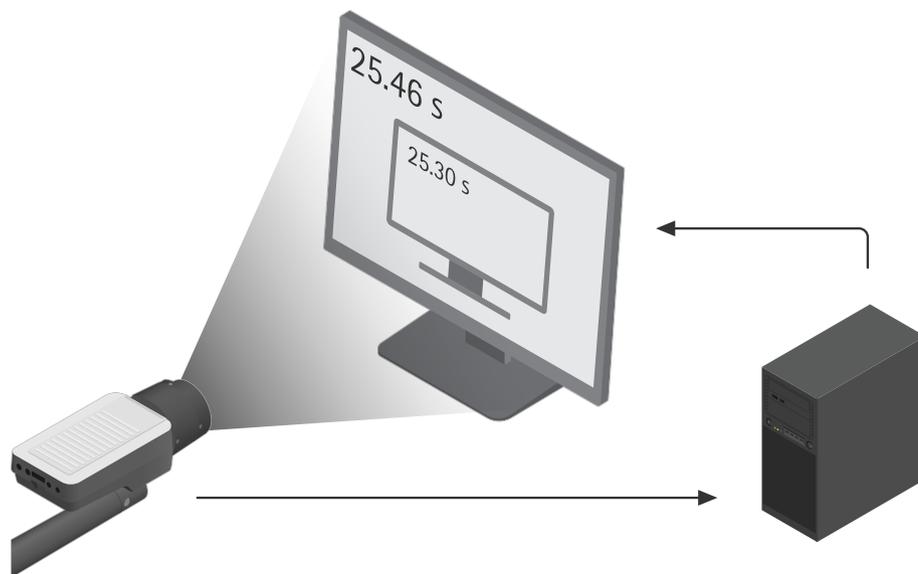
延迟通常以时间单位表示，一般为秒或毫秒 (ms)。精确测量延迟非常困难，因为这需要摄像机和显示设备的时钟精确同步。有一种简单的方法（保留与准确值的最小偏差）：使用时间戳叠加文本功能。这种方法测量的是视频监控系统的端到端延迟，即从镜头捕捉一帧图像到同一帧图像在监视设备上渲染之间的时间差。

请注意，这种方法可能会产生最多一个帧间隔的误差。这取决于用于计算延迟的时间戳只在帧捕获时收集。因此，我们只能用帧速的系数来计算延迟。所以，如果帧速为25 fps，我们可以将延迟计算为40 ms的倍数。如果帧速为1 fps，我们可以将延迟计算为秒的倍数。因此，在帧速较低的情况下，不建议采用这种方法。

使用时间戳叠加文本功能测量延迟：

1. 使用 (%T:%f) 打开叠加中的时间戳
2. 将摄像机放置一定角度，以便其取景自己的实时流输出

3. 拍摄直播流输出的快照，对比原始文本叠加中显示的时间与屏幕循环中显示的时间的时差



使用时间戳叠加文本功能测量延迟的示例。我们可以发现时间差为 $25.46 - 25.30 = 0.16$ 秒，即160 ms。这意味着端到端延迟为160 ms。

4 什么会影响延迟？

总延迟是摄像机、网络 and 客户端带来的延迟之和。

4.1 摄像机引起的延迟

每帧图像的曝光时间间隔大约为 $1/30$ s，然后是短时间的图像缩放和编码。编码后的图像被切块打包，每33 ms向网络输出一幅图像。这一过程在摄像机中花费的时间可以少于50 ms，但更常见的是几百毫秒。根据摄像机（PTZ除外）以及帧是I帧还是P帧的不同，该时间略有差异。

4.1.1 捕捉延迟

让我们来看看视频摄像机的内部结构。摄像机的图像传感器由数百万个光电探测器（光敏点）组成，称为像素。传感器在一次曝光间隔内捕捉像素中的光，然后将记录的光能转换成电子信号。然后像素清空，准备再次曝光。传感器在每个时间单位内的曝光次数（即摄像机每秒能捕捉多少帧）决定传感器的捕捉率。

捕捉延迟取决于捕捉率。如果将捕捉率设置为30 fps，即传感器每 $1/30$ 秒捕捉一张图像，那么捕捉延迟将高达33.3 ms。

4.1.2 图像处理延迟

每个捕捉到的图像帧都要经过去隔行扫描、缩放和图像旋转等图像处理步骤管道，这些步骤会增加延迟。处理数量越多，摄像机带来的延迟就越大。但由于摄像机中的处理过程会影响生成的数据量，因此当数据通过网络传输时，处理过程也会导致对网络延迟产生影响。

影响延迟的一些参数包括：

- **图像旋转。**将视频流旋转90度或270度会给编码处理器增加额外负担。像素需要重新排列和缓冲，这会造成一定的延迟。
- **分辨率。**更高的分辨率意味着处理器需要编码更多的像素。与低分辨率相比，高分辨率所增加的处理时间通常微不足道，因为高分辨率摄像机的处理单元速度更快，可以平衡这一影响。但分辨率越高，每帧的数据量也越大，因此需要传输的数据包也越多。在带宽有限的网络中，这可能会导致传输延迟。这又会导致接收端需要更大的缓冲区，造成更长时间的延迟。
- **噪声过滤。**先进噪声过滤技术需要缓冲多个帧，这会带来额外的延迟。
- **隐私遮罩。**AXIS Live Privacy Shield等先进隐私遮罩功能会带来额外的延迟。这是因为需要进行缓冲，以确保在正确的时间应用正确的隐私遮罩。

4.1.3 压缩延迟

视频需要编码，以便在传输之前压缩数据。压缩涉及一种或几种删除图像数据的数学算法。这需要或多或少的时间，取决于需要处理的数据量。这一步骤中带来的压缩延迟受几个方面的影响：

- **压缩算法的复杂程度**

H.264和H.265是比MJPEG更先进的压缩标准。不过，与MJPEG压缩相比，安讯士摄像机通常具有更高的H.264和H.265压缩容量，这意味着使用H.264或H.265压缩比标准时的延迟不一定更高。另一方面，解码端的延迟可能更高。安讯士摄像机生成的H.264和H.265数据流要求解码器至少缓冲一帧，而MJPEG解码则不需要缓冲。此外，Zipstream存储配置文件最多可增加两帧的额外延迟，相当于30 fps视频延迟66.7 ms。

- **压缩方法的有效性**

安讯士摄像机中常用的编码方案是MJPEG、H.264和H.265。它们都会在摄像机中带来延迟。H.264和H.265比MJPEG更能减少视频吞吐量，这意味着通过网络发送、解包和在接收端渲染的数据包更少。这将减少总延迟。

- **比特率选择**

视频压缩可减小视频数据的大小。然而，压缩后并不是所有帧的大小都保持相同。压缩数据的大小因场景而异。换句话说，原始压缩数据由可变比特率 (VBR) 的流组成，从而导致可变比特率输出到网络中。您需要考虑可用网络的限制，如带宽限制。流式视频系统的带宽限制通常要求对传输比特率进行调节。在某些编码器中，您可以在VBR与最大比特率 (MBR) 之间进行选择。选择MBR可以保证网络接收的数据量有限。通过避免网络超载，可以减少网络延迟以及系统接收端对较大缓冲区的需求。

在安讯士摄像机中，H.264和H.265编码器提供VBR和MBR两种选择。不过，我们通常建议在网络视频中使用VBR，因为质量会根据场景内容实时调整。不建议始终将MBR用作一般的存储缩减工具或用来修复薄弱的网络连接，因为在临界情况下，传输MBR视频的摄像机可能会被强制删除重要的取证细节。

在摄像机中使用Zipstream可降低比特率。这可以保持数据量的减少，从而减少网络延迟。

- **流的个数。**如果要求摄像机提供一种以上的流类型（不同的帧速或分辨率），处理额外的流类型时将增加延迟，因为各个流必须由同一个处理器编码。

在选择压缩方式时，您应考虑到这些方面。一方面，先进的编码算法可能需要较长的编码和解码时间，但另一方面，它将减少通过互联网发送的数据量，进而缩短传输延迟，减少接收器缓冲区的大小。

4.1.4 缓冲延迟

由于视频是一帧一帧处理的，因此一次只能压缩有限的的数据量。在各处理阶段之间有时需要短期缓冲，这就造成了摄像机中的延迟。

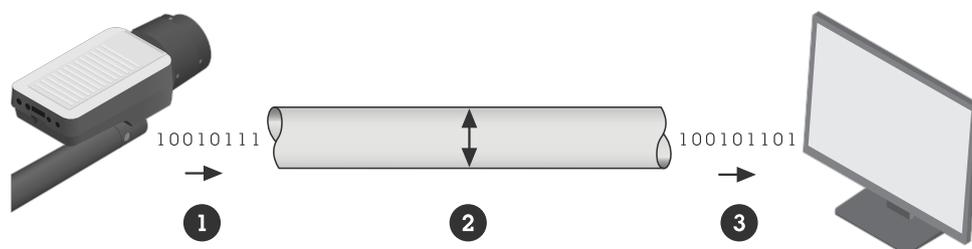
4.1.5 音频延迟

在某些情况下，视频流还伴有音频。音频编码器需要等待一定的采样数量，然后才能有可用数据块，以便开始音频编码，这会增加摄像机端的延迟。不同音频编码算法的采样率和数据块大小不同。

4.2 网络延迟

在捕捉、处理和压缩图像之后，视频数据将通过网络传输，然后到达客户端进行渲染。为了了解网络如何影响延迟，我们首先需要了解视频网络中的一些基本概念：比特率、带宽和吞吐量。网络延迟与比特率成正比，与带宽成反比。

- **比特率**是指单位时间内处理的数据量，以“比特”为单位。在视频监控中，比特率是指单位时间内通过网络发送的摄像机生成的数据量。比特率取决于很多因素；它在很大程度上取决于拍摄的场景、摄像机所做的处理以及视频流设置。当摄像机生成更多数据需要传输时，如果带宽有限，就会出现较高的网络延迟。
- **带宽**是指摄像机与监视器之间的网络可以处理的数据量。这是指链路的最大传输能力。带宽取决于链路的长度和基础设施，即交换机、路由器、电缆、代理服务器等。如果我们增加网络容量，就能够通过更多数据，从而降低延迟。
- **吞吐量**是实际实现的数据传输速度（比特/秒）。这取决于您是否与他人共用链路，以及链路中电缆的电磁干扰。端口上配置的QoS（服务质量）设置也可能对吞吐量设置上限。



如果我们把摄像机与监视器之间的链路（网络）比作一根管道，比特率(1)就是单位时间内传输到管道内的数据量，带宽(2)就是管道的粗细，吞吐量(3)就是单位时间内实际通过管道的数据量。

- 1 比特率
- 2 带宽
- 3 吞吐量

网络中的总延迟取决于三大因素：摄像机与视频查看设备之间链路的基础设施（决定带宽）、摄像机生成的数据量（决定比特率）以及传输协议的选择。

4.2.1 基础设施

网络是非常不可预测的端到端延迟来源。交换机、路由器、电缆、代理服务器：发送端与接收端之间网络中的各个方面都会影响总延迟。在局域网 (LAN) 中，网络延迟可能只有几毫

秒，微不足道，可以忽略不计。但是，如果视频流要在互联网上以未指定路由的方式传输，网络延迟将很难预测，而且在很多情况下可能是造成端到端延迟的主要原因。

通过精心的网络管理和带宽分配，网络延迟的不可预测因素可以变得更加可预测。摄像机和查看设备之间的链路需要有保证的吞吐量。

在局域网 (LAN) 中，这可以通过确保链路中的跃点尽可能少来实现。链路不得与其他流量共享，例如默认情况下优先于视频的互联网协议电话 (VoIP) 或其他协议，或其他可能使链路超载的服务。

如果链路是通过广域网 (WAN) 实现的，则需要每个跃点（即路由器和交换机）中保证服务质量 (QoS)。这也可以通过在本地互联网提供商处租用点对点路由来实现。

影响吞吐量的可配置因素：

- 数据包开销（取决于协议，例如：VLAN标头）
- 代理服务器和防火墙
- 整个路由中每个链路的QoS
- 是否启用突发模式（启用 → 更高速度）
- MTU – 视频有效载荷的大小

影响吞吐量的成本相关因素：

- 交换机和路由器的处理器速度以及端口缓冲
- 电缆类型（或无线）

4.2.2 视频流数据量

摄像机中图像处理和压缩方法的选择会影响网络延迟，因为这些选择影响生成的视频数据量。发送少量数据所需的时间显然更短。

4.2.3 传输协议

来自摄像机的视频帧传送到传输协议应用（通常是RTP或HTTP），然后通过互联网协议 (IP) 网络传送到渲染客户端。传输可以通过可靠的TCP（一种面向连接的协议，对丢失的数据包进行重传）进行，或者通过UDP（一种较简单的协议，不保证交付，也不提供丢失数据包重传的设施）进行。

对于安讯士摄像机，在封装编码数据流以进行传输时有以下选项：

表 4.1 封装编码数据流的选项。

拓扑	推荐的安讯士视频数据包封装模式
局域网 (LAN)/少跃点和直接管理节点	MJPEG / HTTP / TCP
局域网 (LAN)/少跃点和直接管理节点	H.264、H.265或MJPEG / RTP / RTSP / HTTP / TCP
局域网 (LAN)/少跃点和直接管理节点	H.264、H.265或MJPEG / RTP / RTSP / TCP
广域网 (WAN)/多跃点，您无法完全控制节点	H.264、H.265或MJPEG / RTP / 单播 / UDP

表 4.1. 封装编码数据流的选项。(续)

拓扑	推荐的安讯士视频数据包封装模式
广域网 (WAN)/多跃点, 您无法完全控制节点	H.264、H.265或MJPEG / RTP / 组播 / UDP
远程连接 / 云 / 广域网 (WAN)/多跃点, 您无法完全控制节点	H.264、H.265或MJPEG / RTP / WebRTC / UDP或TCP

通常情况下, 通过TCP传输一个数据包比通过UDP传输一个数据包需要更长的时间, 因为需要额外的连接设置、确认信息以及在侦测到数据包丢失时重新传输。另一方面, 使用UDP时, 当数据包在途中丢失时, 用户将在视频流中遇到伪影或中断。TCP会在数据包丢失时产生抖动, 而UDP则会在数据包丢失时产生伪影和/或中断。如果数据丢失和暂时的质量下降可以接受, 那么UDP可能是带宽较低的网络的一种选择。

如果使用TCP, 需要发送的数据包更多, 同时, 为了支持这种情况, 需要更大的带宽。如果知道网络中存在大量拥塞, 则应选择UDP作为传输协议。因为接受丢包, 同时它也会导致丢包, 从而降低图像质量。

通过WebRTC和自适应比特率, 视频根据网络进行调节, 避免使用TCP时可能出现的无法控制的延迟峰值。

4.3 客户端延迟

视频系统客户端接收视频后, 会对其进行解包、重新排序和解码, 并使用媒体播放器来渲染视频。每个步骤都会造成客户端总延迟增加, 而计算机本身也会产生重要影响。CPU容量、操作系统、网卡和显卡都会影响延迟。通常, MJPEG方法的解码和显示延迟更低, 因为数据可以在到达时在屏幕上呈现, 无需时间码。H.264和其他视频压缩标准为每张图片分配时间码, 并要求它们得到相应渲染。

4.3.1 播放缓冲

真实的网络往往非常庞大和复杂, 存在突发流量情况和以不同顺序到达的数据包。为了补偿网络传输带来的变化, 客户端使用缓冲区。该缓冲区称为播放缓冲区或抖动缓冲区, 可确保数据包按正确的顺序排列, 并缓冲足够的帧, 使解码器不会处于“饥饿”状态, 从而在查看器中显示统一的帧速。

该缓冲区可造成相对较高的延迟。不同的查看器应用软件有不同的播放缓冲区大小, 在大多数查看器中, 您可以对其进行更改。但需要注意的是, 减少缓冲区会增加抖动, 因此需要在抖动与可接受延迟之间找到平衡点。

4.3.2 音频缓冲

在回放过程中, 音频流比视频流对卡顿或延迟更敏感。一个延迟的音频数据包可在声道中产生恼人的空白, 而且音频必须与视频唇形同步。因此, 当视频伴有音频时, 需要设置较大的播放缓冲区, 这自然也会增加延迟。

4.3.3 解压

解压缩过程所需的时间取决于所使用的编码格式。解码延迟在很大程度上取决于显卡提供的硬件解码器支持。硬件解码通常比软件解码快。一般来说, H.264解码比MJPEG更耗时。对于H.264解码, 延迟还取决于您在编码阶段选择的配置文件。基础水平最容易解码, 而主要水平和高水平则需要较长时间。安讯士视频产品生成的H.264数据流要求解码器至少缓冲一个帧。

4.3.4 显示设备

显示设备也会通过传输时间、刷新率和响应时间影响延迟。

传输时间是指解码后的视频数据从解码器通过电缆（例如HDMI）发送到监视器所需的时间。传输时间取决于电缆的速度和监视器的分辨率。对于使用标准HDMI电缆连接的全高清监视器，这会增加约10 ms的延迟。

显示设备的刷新率也会影响延迟。电脑监视器的刷新率约为17–20 ms，而专用游戏监视器的刷新率为4–5 ms。

响应时间是指监视器中的像素改变数值所需的时间。这取决于变化的大小，但对于较大的变化，可能会增加5–20 ms的延迟。

5 减少延迟

设计满足低延迟目标的系统还需要进行其他平衡。您需要决定可接受的延迟时间，并在视频质量与监控系统成本之间找到平衡点。本章提供了一些关于摄像机端、网络和客户端的简单建议，以减少端到端延迟。

5.1 减少摄像机延迟

- **分辨率。** 尽可能选择较低的分辨率。更高的分辨率意味着需要编码更多的数据，从而可能导致更高的延迟。
- **增强功能。** 减少图像增强，如旋转、去隔行扫描和缩放。使用图像增强功能会增加延迟。
- **低延迟模式。** 通过在普通配置设置中开启低延迟模式，您可以优化实时流的图像处理时间。直播流的延迟降到更低，但图像质量也会比平时更低。
- **隐私遮罩。** 考虑不使用隐私遮罩，因为它会增加延迟。
- **编码。** 确保编码器能提供系统所需的延迟控制水平。需要在数据量与网络基础设施容量之间保持平衡。如果通过带宽有限的网络发送视频，选择H.264或H.265作为编码格式。这将导致压缩比更大，从而降低比特率。如果网络可以管理比特率，选择baseline profile，因为baseline更容易编码和解码。
- **Zipstream中的存储配置文件。** 考虑不使用存储配置文件，因为它会增加延迟。
- **流的个数。** 用不同的设置来限制摄像机的流个数。每种独特的设置组合（如分辨率、帧速和压缩率）都需要单独的编码过程，这会增加处理器的负荷并导致延迟。
- **比特率。** 尽量使用较低的比特率。为了减少摄像机的延迟，我们需要减少生成的数据量。比特率受很多因素影响，如光线条件、场景类型以及压缩水平、分辨率、帧速等。
- **帧速。** 尽可能使用高帧速。由于帧的编码和解码都是一帧一帧进行的，因此缓冲至少会延迟一帧。帧速越高，缓冲造成的延迟就越小。对于30 fps的视频流，每帧的捕捉时间为1/30秒。这样，我们就可以预计缓冲的延迟高达33 ms。如果是25 fps，延迟将达到40 ms。
- **取景模式。** 考虑使用分辨率尽可能低、帧速尽可能高的取景模式。低分辨率意味着需要处理的像素更少，高帧速意味着缓冲延迟更低。
- **音频。** 考虑不使用音频。音频必须与视频同步，需要更大的回放缓冲区，从而导致更高的延迟。

5.2 减少网络延迟

有关摄像机端的许多建议都旨在限制通过网络发送的总数据量。在大多数情况下，有限的网络是造成端到端延迟的主要因素。如果网络容量大，则不需要遵循上述许多建议。确保您的网络具有良好的服务质量，网络内的跃点都经过适当配置，以满足您的视频需求。确保网络比特率能够传输摄像机输出的数据。

5.3 减少客户端延迟

客户端改进对端到端延迟有很大影响，通常可以进行很多操作。

处理器和显卡。 CPU在客户端延迟中起着核心作用。使用性能良好的处理器，其容量足以处理视频流并同时处理其他请求。使用性能良好的显卡，并配备更新的软件和解码支持。

查看器/VMS。 确保您的查看器没有不必要的过长播放缓冲，但要注意与视频抖动的平衡。

显示屏。 使用刷新率尽可能快的显示屏。为了获得满意的Live View（实时画面）（而不影响延迟），将屏幕频率调整为摄像机捕捉帧速的倍数。例如，30 fps模式设置为60 Hz，25 fps模式设置为50 Hz。

关于 Axis Communications

Axis 通过打造解决方案，不断提供改善以提高安全性和业务绩效。作为网络技术公司和行业领导者，Axis 提供视频监控解决方案，访问控制、对讲以及音频系统的相关产品和服务。并通过智能分析应用实现增强，通过高品质培训提供支持。

Axis 在 50 多个国家/地区拥有约 4,000 名敬业的员工 并与全球的技术和系统集成合作伙伴合作 为客户带来解决方案。Axis 成立于 1984 年，总部在瑞典隆德