

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 1 月 16 日 (16.01.2020)



(10) 国际公布号
WO 2020/010709 A1

(51) 国际专利分类号:
H04N 19/40 (2014.01)

中国上海市徐汇区斜土路2899号甲光启文化
广场A幢5楼, Shanghai 200030 (CN)。

(21) 国际申请号: PCT/CN2018/107476

(22) 国际申请日: 2018 年 9 月 26 日 (26.09.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810770745.4 2018年7月13日 (13.07.2018) CN

(71) 申请人: 网宿科技股份有限公司 (WANGSU
SCIENCE & TECHNOLOGY CO., LTD.) [CN/CN];
中国上海市徐汇区斜土路2899号甲光启文化
广场A幢5楼, Shanghai 200030 (CN)。

(72) 发明人: 郑惠洁(ZHENG, Huijie); 中国上海市徐汇
区斜土路2899号甲光启文化广场A幢5楼, Shanghai
200030 (CN)。 刘珠清(LIU, Zhuqing); 中国上海
市徐汇区斜土路2899号甲光启文化广场A幢5
楼, Shanghai 200030 (CN)。 李俊明(LI, Junming);

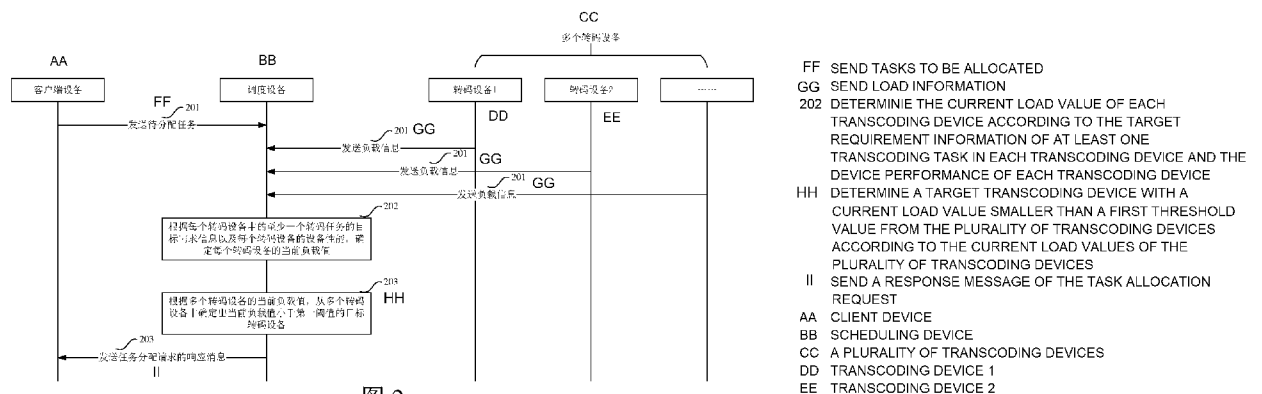
(74) 代理人: 北京华智则铭知识产权代理有
限公司(BEIJING HUA-WISDOM INTELLECTUAL
PROPERTY INC.); 中国北京市大兴区宏业路9号
院7号楼1307室, Beijing 100162 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家
保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG,
BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU,
CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB,
GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,
JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK,
LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,
MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,
PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,

(54) Title: A TRANSCODING TASK ALLOCATION METHOD, SCHEDULING DEVICE AND TRANSCODING DEVICE

(54) 发明名称: 一种转码任务的分配方法及调度设备、转码设备



(57) Abstract: Disclosed by the invention are a transcoding task allocation method, a scheduling device and a transcoding device. The method comprises: once the scheduling device receives a task allocation request sent by a client device, determining the current load value of each transcoding device according to the target resolution of the transcoding tasks carried in each transcoding device and the device performance of each transcoding device, and then determining the target transcoding device with a current load value smaller than a first threshold value, so that the client device can send the transcoding tasks to the target transcoding device. Thereby, compared with the prior art that the target transcoding device is directly selected according to the IP address of the transcoding device, the embodiment of the invention fully takes the load condition of the transcoding device into consideration, effectively avoids the situation that the transcoding task cannot be executed timely due to overload of the transcoding device, and improves the video transcoding efficiency.

NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布：

— 包括国际检索报告(条约第21条(3))。

(57) 摘要：本发明公开了一种转码任务的分配方法及调度设备、转码设备。所述方法包括：调度设备接收到客户端设备发送的任务分配请求后，可以根据每个转码设备中承载的转码任务的目标分辨率以及每个转码设备的设备性能，确定每个转码设备的当前负载值，进而确定当前负载值小于第一阈值的目标转码设备，从而使得客户端设备可以向目标转码设备发送转码任务。如此，相比于现有技术中直接根据转码设备的IP地址来选择目标转码设备的方式而言，本发明实施例充分考虑了转码设备的负载情况，有效避免了因转码设备过载而不能及时执行转码任务的情况，提高了视频转码的效率。

一种转码任务的分配方法及调度设备、转码设备

技术领域

本发明涉及数据处理领域，尤其涉及一种转码任务的分配方法及调度设备、转码设备。

背景技术

在视频网站中，视频转码是视频生成过程中的重要环节。视频网站将丰富的视频资源转换成适合于网络播放的格式，从而，用户便可以通过计算机终端、移动终端或者具有联网功能的电视终端访问相应的视频网站观看各类视频。由于不同的视频资源可能具有不同的视频格式，为了适配不同网络带宽的场景，需要将视频以不同清晰度呈现给用户。因此，视频网站需要建立转码任务，通过其内部的转码系统执行相应的转码任务进行视频转码处理。

现有的转码系统中，调度设备在接收到客户端设备发送的转码任务后，通常根据客户端设备的网络之间互连的协议（Internet Protocol，IP）地址选择合适的转码设备，以实现客户端设备和该转码设备之间执行转码任务。然而，采用这种方法选出的转码设备可能存在过载的情况，从而导致该转码设备不能及时执行客户端设备发送的转码任务。

基于此，目前亟需一种转码任务的分配方法，用于解决现有技术分配的转码设备由于过载等原因不能及时执行转码任务的问题。

发明内容

本发明实施例提供一种转码任务的分配方法及调度设备、转码设备，以解决现有技术分配的转码设备由于过载等原因不能及时执行转码任务的技术问题。

本发明实施例提供一种转码任务的分配方法，所述方法包括：

调度设备接收到客户端设备发送的任务分配请求后，获取多个转码设备的负载信息，每个转码设备的负载信息包括转码设备承载的转码任务的目标需求信息，所述转码任务的目标需求信息包括所述转码任务的目标分辨率和所述转码任务的目标帧率；

针对第一转码设备，所述调度设备根据所述第一转码设备中的至少一个转码任务的目标需求信息以及所述第一转码设备的设备性能，确定所述第一转码设备的当前负载值，所述第一转码设备为所述多个转码设备中的任一设备；

所述调度设备根据所述多个转码设备的当前负载值，从所述多个转码设备中确定出当前负载值小于第一阈值的目标转码设备，并向所述客户端设备返回所述任务分配请求的响应消息，所述响应消息用于指示所述客户端设备向所述目标转码设备发送转码任务。

如此，调度设备可以根据每个转码设备的负载信息来计算每个转码设备的负载值，从而可以选择负载值较小的转码设备作为目标转码设备，相比于现有技术中直接根据转码设备的 IP 地址来选择目标转码设备的方式而言，本发明实施例充分考虑了转码设备的负载情况，有效避免了因转码设备过载而不能及时执行转码任务的情况，提高了视频转码的效率。

在一种可能的实现方式中，所述调度设备根据所述第一转码设备中的至少一个转码任务的目标需求信息以及所述第一转码设备的设备性能，确定所述第一转码设备的当前负载值，包括：

所述调度设备根据所述第一转码设备的设备性能，确定所述第一转码设备承载的各转码任务的各目标需求信息对应的预设最大任务数量；

所述调度设备根据所述第一转码设备中每个目标需求信息对应的预设最大任务数量，以及所述第一转码设备中每个目标需求信息对应的转码任务的数量，确定所述第一转码设备的当前负载值。

如此，调度设备可以根据各目标需求信息在第一转码设备中的任务数量以

及预设最大任务数量，得到第一转码设备的当前负载值，这种方法能够更加合理地计算转码设备的当前负载值，从而提高确定转码设备的合理性。

在一种可能的实现方式中，所述当前负载值通过以下方式确定：

$$A=f_1\times\frac{1}{X_1}+f_2\times\frac{1}{X_2}+\cdots+f_n\times\frac{1}{X_n}$$

其中，A 为所述第一转码设备的当前负载值； f_i 为所述第一转码设备中第 i 个目标需求信息对应的任务数量， X_i 为所述第一转码设备中第 i 个目标需求信息对应的预设最大任务数量， $i=1, 2, \dots, n$ ；n 为所述第一转码设备承载的转码任务的目标需求信息的数目，n 为大于或等于 1 的整数。

在一种可能的实现方式中，所述每个转码设备的负载信息还包括所述每个转码设备中多个执行核心的使用率；

所述调度设备根据所述多个转码设备的当前负载值，从所述多个转码设备中确定出当前负载值小于第一阈值的目标转码设备，包括：

所述调度设备根据所述多个转码设备的当前负载值和所述多个转码设备的平均中央处理器CPU使用率，从所述多个转码设备中确定出当前负载值小于第一阈值，且平均CPU使用率小于第二阈值的目标转码设备；所述转码设备的平均CPU使用率是根据转码设备中的多个执行核心的使用率确定的。

如此，在确定目标转码设备时，既可以考虑每个转码设备的当前负载值，又可以考虑每个转码设备的平均CPU使用率，能够从不同的角度来选择转码设备，提高确定转码设备的合理性。

在一种可能的实现方式中，所述调度设备根据所述第一转码设备中的多个执行核心的使用率，确定所述第一转码设备的平均CPU使用率，包括：

所述调度设备将所述第一转码设备中的多个执行核心按照每个执行核心的使用率从大到小进行排名；

所述调度设备将排名前N位的执行核心的使用率的平均值确定为所述第一转码设备的平均CPU使用率，N为大于或等于1的整数。

如此，能够避免转码设备中执行核心过多时计算量大的问题，从而降低平

均CPU使用率的计算量，进而降低调度设备的运行负担。

在一种可能的实现方式中，所述从所述多个转码设备中确定出当前负载值小于第一阈值，且平均CPU使用率小于第二阈值的目标转码设备，包括：

所述调度设备从所述多个转码设备中确定出当前负载值小于所述第一阈值的至少一个候选转码设备；

所述调度设备根据每个候选转码设备中的多个执行核心的使用率，确定所述每个候选转码设备的平均CPU使用率；

所述调度设备从所述至少一个候选转码设备中确定出平均CPU使用率小于所述第二阈值的目标转码设备。

在一种可能的实现方式中，所述任务分配请求包括待分配任务的目标需求信息，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率和所述待分配任务的目标帧率；

所述调度设备根据所述多个转码设备的当前负载值，从所述多个转码设备中确定出当前负载值小于第一阈值的目标转码设备，包括：

所述调度设备根据所述多个转码设备的当前负载值和所述多个转码设备的预测负载值，从所述多个转码设备中确定出当前负载值小于第一阈值，且预测负载值小于第三阈值的目标转码设备；所述转码设备的预设负载值是根据所述转码设备的当前负载值和所述待分配转码任务的负载值确定的，所述待分配任务的负载值是根据所述待分配转码任务的目标需求信息确定的。

如此，在确定目标转码设备时，既可以考虑每个转码设备的当前负载值，又可以考虑每个转码设备的预测负载值，能够从不同的角度来选择转码设备，提高确定转码设备的合理性。

在一种可能的实现方式中，所述任务分配请求包括待分配任务的目标需求信息和所述待分配任务的标识，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率、所述待分配任务的目标帧率；所述每个转码设备的负载信息还包括转码设备承载的转码任务的标识；

在所述获取多个转码设备的负载信息之后，在所述调度设备确定所述第一转码设备的当前负载值之前，所述方法还包括：

所述调度设备确定所述多个转码设备承载的转码任务中不存在与所述待分配任务的目标分辨率、所述待分配任务的目标帧率和所述待分配任务的标识均一致的转码任务。

如此，调度设备能够优先判断多个转码设备中是否已存在与待分配任务一致的转码任务，从而避免在待分配任务已存在的情况下重复创建转码任务，导致资源浪费的情况。

本发明实施例提供一种转码任务的分配方法，所述方法包括：

转码设备接收到调度设备发送的状态请求后，根据自身的负载信息以及自身的设备性能，确定当前负载值；所述负载信息包括所述转码设备承载的至少一个转码任务的目标需求信息，所述转码任务的目标需求信息包括所述转码任务的目标分辨率和所述转码任务的目标帧率；

所述转码设备根据所述当前负载值，生成所述状态请求的响应消息；

所述转码设备向所述调度设备发送的所述响应消息，以使所述调度设备根据所述响应消息确定出目标转码设备，所述目标转码设备用于执行待分配的转码任务。

如此，转码设备可以根据自身的负载信息来计算当前负载值，可以减轻调度设备的计算量，从而减轻调度设备的运行负担；进一步地，调度设备可以选择当前负载值较小的转码设备作为目标转码设备，相比于现有技术中直接根据转码设备的IP地址来选择目标转码设备的方式而言，本发明实施例充分考虑了转码设备的负载情况，有效避免了因转码设备过载而不能及时执行转码任务的情况，提高了视频转码的效率。

在一种可能的实现方式中，所述转码设备根据自身的负载信息以及自身的设备性能，确定当前负载值，包括：

所述转码设备根据所述设备性能，确定所承载的各转码任务的各目标需求

信息对应的预设最大任务数量；

所述转码设备根据每个目标需求信息对应的预设最大任务数量，以及所述每个目标需求信息对应的转码任务的数量，确定所述转码设备的当前负载值。

如此，转码设备可以根据各目标需求信息在该转码设备中的任务数量以及预设最大任务数量，得到当前负载值，这种方法能够更加合理地计算转码设备的当前负载值，从而提高确定目标转码设备的合理性。

在一种可能的实现方式中，所述当前负载值通过以下方式确定：

$$A=f_1\times\frac{1}{X_1}+f_2\times\frac{1}{X_2}+\cdots+f_n\times\frac{1}{X_n}$$

其中，A 为所述转码设备的当前负载值； f_i 为所述转码设备中第 i 个目标需求信息对应的任务数量， X_i 为所述转码设备中第 i 个目标需求信息对应的预设最大任务数量， $i=1, 2, \dots, n$ ；n 为所述转码设备承载的转码任务的目标需求信息的数目，n 为大于或等于 1 的整数。

在一种可能的实现方式中，所述负载信息还包括多个执行核心的使用率；所述方法还包括：

所述转码设备根据所述多个执行核心的使用率，确定平均CPU使用率；

所述转码设备根据所述当前负载值，生成所述状态请求的响应消息，包括：

所述转码设备根据所述当前负载值和所述确定平均CPU使用率，生成所述状态请求的响应消息。

在一种可能的实现方式中，所述转码设备根据所述多个执行核心的使用率，确定平均CPU使用率，包括：

所述转码设备将所述多个执行核心按照每个执行核心的使用率从大到小进行排名；

所述转码设备将排名前N位的执行核心的使用率的平均值确定为所述平均CPU使用率，N为大于或等于1的整数。

如此，能够避免转码设备中执行核心过多时计算量大的问题，从而降低平均CPU使用率的计算量，进而降低转码设备的运行负担。

在一种可能的实现方式中，所述状态请求包括所述待分配任务的目标需求信息，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率和所述待分配任务的目标帧率；

所述方法还包括：

所述转码设备根据所述待分配任务的目标需求信息，确定所述待分配任务的负载值；

所述转码设备根据所述当前负载值和所述待分配任务的负载值，确定自身的预测负载值；

所述转码设备根据所述当前负载值，生成所述状态请求的响应消息，包括：

所述转码设备根据所述当前负载值和所述预测负载值，生成所述状态请求的响应消息。

在一种可能的实现方式中，所述状态请求包括所述待分配任务的目标需求信息和所述待分配任务的标识，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率、所述待分配任务的目标帧率；

转码设备在接收到调度设备发送的状态请求之后，在确定所述当前负载值之前，所述方法还包括：

所述转码设备确定所述转码设备承载的转码任务中不存在与所述待分配任务的目标分辨率、所述待分配任务的目标帧率和所述待分配任务的标识均一致的转码任务。

如此，转码设备能够优先判断是否已存在与待分配任务一致的转码任务，从而避免在待分配任务已存在的情况下重复创建转码任务，导致资源浪费的情况。

本发明实施例提供一种调度设备，所述调度设备包括：

接收单元，用于接收到客户端设备发送的任务分配请求后，获取多个转码设备的负载信息，每个转码设备的负载信息包括转码设备承载的转码任务的目标需求信息，所述转码任务的目标需求信息包括所述转码任务的目标分辨率和

所述转码任务的目标帧率；

处理单元，用于针对第一转码设备，根据所述第一转码设备中的至少一个转码任务的目标需求信息以及所述第一转码设备的设备性能，确定所述第一转码设备的当前负载值，所述第一转码设备为所述多个转码设备中的任一设备；以及根据所述多个转码设备的当前负载值，从所述多个转码设备中确定出当前负载值小于第一阈值的目标转码设备；

发送单元，用于向所述客户端设备返回所述任务分配请求的响应消息，所述响应消息用于指示所述客户端设备向所述目标转码设备发送转码任务。

在一种可能的实现方式中，所述处理单元具体用于：

根据所述第一转码设备的设备性能，确定所述第一转码设备承载的各转码任务的各目标需求信息对应的预设最大任务数量；

根据所述第一转码设备中每个目标需求信息对应的预设最大任务数量，以及所述第一转码设备中每个目标需求信息对应的转码任务的数量，确定所述第一转码设备的当前负载值。

在一种可能的实现方式中，所述当前负载值通过以下方式确定：

$$A=f_1\times\frac{1}{X_1}+f_2\times\frac{1}{X_2}+\cdots+f_n\times\frac{1}{X_n}$$

其中，A 为所述第一转码设备的当前负载值； f_i 为所述第一转码设备中第 i 个目标需求信息对应的任务数量， X_i 为所述第一转码设备中第 i 个目标需求信息对应的预设最大任务数量， $i=1, 2, \dots, n$ ；n 为所述第一转码设备承载的转码任务的目标需求信息的数目，n 为大于或等于 1 的整数。

在一种可能的实现方式中，所述每个转码设备的负载信息还包括所述每个转码设备中多个执行核心的使用率；

所述处理单元具体用于：

根据所述多个转码设备的当前负载值和所述多个转码设备的平均中央处理器CPU使用率，从所述多个转码设备中确定出当前负载值小于第一阈值，且平均CPU使用率小于第二阈值的目标转码设备；所述转码设备的平均CPU使用

率是根据转码设备中的多个执行核心的使用率确定的。

在一种可能的实现方式中，所述处理单元具体用于：

将所述第一转码设备中的多个执行核心按照每个执行核心的使用率从大到小进行排名；

将排名前N位的执行核心的使用率的平均值确定为所述第一转码设备的平均CPU使用率，N为大于或等于1的整数。

在一种可能的实现方式中，所述处理单元具体用于：

从所述多个转码设备中确定出当前负载值小于所述第一阈值的至少一个候选转码设备；

根据每个候选转码设备中的多个执行核心的使用率，确定所述每个候选转码设备的平均CPU使用率；

从所述至少一个候选转码设备中确定出平均CPU使用率小于所述第二阈值的值的目标转码设备。

在一种可能的实现方式中，所述任务分配请求包括待分配任务的目标需求信息，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率和所述待分配任务的目标帧率；

所述处理单元具体用于：

根据所述多个转码设备的当前负载值和所述多个转码设备的预测负载值，从所述多个转码设备中确定出当前负载值小于第一阈值，且预测负载值小于第三阈值的目标转码设备；所述转码设备的预设负载值是根据所述转码设备的当前负载值和所述待分配转码任务的负载值确定的，所述待分配任务的负载值是根据所述待分配转码任务的目标需求信息确定的。

在一种可能的实现方式中，所述任务分配请求包括待分配任务的目标需求信息和所述待分配任务的标识，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率、所述待分配任务的目标帧率；所述每个转码设备的负载信息还包括转码设备承载的转码任务的标识；

所述处理单元在确定所述第一转码设备的当前负载值之前，还用于：

确定所述多个转码设备承载的转码任务中不存在与所述待分配任务的目标分辨率、所述待分配任务的目标帧率和所述待分配任务的标识均一致的转码任务。

本发明实施例提供一种转码设备，所述转码设备包括：

接收单元，用于接收到调度设备发送的状态请求；

处理单元，用于根据自身的负载信息以及自身的设备性能，确定当前负载值；所述负载信息包括所述转码设备承载的至少一个转码任务的目标需求信息，所述转码任务的目标需求信息包括所述转码任务的目标分辨率和所述转码任务的目标帧率；以及根据所述当前负载值，生成所述状态请求的响应消息；

发送单元，用于向所述调度设备发送的所述响应消息，以使所述调度设备根据所述响应消息确定出目标转码设备，所述目标转码设备用于执行待分配的转码任务。

在一种可能的实现方式中，所述处理单元具体用于：

根据所述设备性能，确定所承载的各转码任务的各目标需求信息对应的预设最大任务数量；

根据每个目标需求信息对应的预设最大任务数量，以及所述每个目标需求信息对应的转码任务的数量，确定所述转码设备的当前负载值。

在一种可能的实现方式中，所述当前负载值通过以下方式确定：

$$A=f_1\times\frac{1}{X_1}+f_2\times\frac{1}{X_2}+\cdots+f_n\times\frac{1}{X_n}$$

其中，A 为所述转码设备的当前负载值； f_i 为所述转码设备中第 i 个目标需求信息对应的任务数量， X_i 为所述转码设备中第 i 个目标需求信息对应的预设最大任务数量， $i=1, 2, \dots, n$ ；n 为所述转码设备承载的转码任务的目标需求信息的数目，n 为大于或等于 1 的整数。

在一种可能的实现方式中，所述负载信息还包括多个执行核心的使用率；所述处理单元还用于：

根据所述多个执行核心的使用率，确定平均CPU使用率；

根据所述当前负载值和所述确定平均CPU使用率，生成所述状态请求的响应消息。

在一种可能的实现方式中，所述处理单元具体用于：

将所述多个执行核心按照每个执行核心的使用率从大到小进行排名；

将排名前N位的执行核心的使用率的平均值确定为所述平均CPU使用率，N为大于或等于1的整数。

在一种可能的实现方式中，所述状态请求包括所述待分配任务的目标需求信息，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率和所述待分配任务的目标帧率；

所述处理单元还用于：

根据所述待分配任务的目标需求信息，确定所述待分配任务的负载值；

根据所述当前负载值和所述待分配任务的负载值，确定自身的预测负载值；

根据所述当前负载值和所述预测负载值，生成所述状态请求的响应消息。

在一种可能的实现方式中，所述状态请求包括所述待分配任务的目标需求信息和所述待分配任务的标识，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率、所述待分配任务的目标帧率；

所述处理单元在确定所述当前负载值之前，所述处理单元还用于：

确定所述转码设备承载的转码任务中不存在与所述待分配任务的目标分辨率、所述待分配任务的目标帧率和所述待分配任务的标识均一致的转码任务。

附图说明

为了更清楚地说明本发明实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简要介绍。

图1为本发明实施例适用的一种系统架构示意图；

图2为本发明实施例提供的一种转码任务的分配方法所对应的流程示意图；

图 3 为本发明实施例提供的另一种转码任务的分配方法所对应的流程示意图；

图 4 为本发明实施例提供的一种调度设备所对应的结构示意图；

图 5 为本发明实施例提供的一种转码设备所对应的结构示意图。

具体实施方式

下面结合说明书附图对本申请进行具体说明，方法实施例中的具体操作方法也可以应用于装置实施例中。

图 1 示例性示出了本发明实施例适用的一种系统架构示意图，如图 1 所示，本发明实施例适用的系统架构 100 包括调度设备 101，至少一个客户端设备，例如图 1 中所示出的客户端设备 1021、客户端设备 1022 和客户端设备 1023，以及多个转码设备，例如图 1 中所示出的转码设备 1031、转码设备 1032 和转码设备 1033。

本发明实施例中，客户端设备可以为多种类型的设备，例如笔记本电脑、智能手机、平板电脑和智能电视等。

实施例一：

基于图 1 所示的系统架构，图 2 示例性示出了本发明实施例提供的一种转码任务的分配方法所对应的流程示意图，包括以下步骤：

步骤 201，调度设备接收到客户端设备发送的任务分配请求后，获取多个转码设备的负载信息。

步骤 202，针对第一转码设备，所述调度设备根据所述第一转码设备中的至少一个转码任务的目标需求信息以及所述第一转码设备的设备性能，确定所述第一转码设备的当前负载值。

步骤 203，所述调度设备根据所述多个转码设备的当前负载值，从所述多个转码设备中确定出当前负载值小于第一阈值的目标转码设备，并向所述客户端设备返回所述任务分配请求的响应消息，所述响应消息用于指示所述客户端

设备向所述目标转码设备发送转码任务。

本发明实施例中，调度设备可以根据每个转码设备的负载信息来计算每个转码设备的当前负载值，从而可以选择当前负载值较小的转码设备作为目标转码设备，相比于现有技术中直接根据转码设备的 IP 地址来选择目标转码设备的方式而言，本发明实施例充分考虑了转码设备的负载情况，有效避免了因转码设备过载而不能及时执行转码任务的情况，提高了视频转码的效率。

在执行步骤 201 之前，图 1 中示出的任一客户端设备可以向调度设备发送任务分配请求，比如，客户端设备 1021 向调度设备 101 发送任务分配请求。该任务分配请求可以用于请求调度设备向客户端设备分配合适的转码设备。

以视频直播的场景为例，视频主播录制视频后可以上传到直播平台，由于直播平台是面向众多用户的，不同用户对视频的清晰度的要求也不同，这就需要直播平台借助转码设备将视频转码，从而满足用户对不同清晰度的要求。本发明实施例中，直播平台对应的设备可以作为客户端设备，直播平台在接收到视频主播上传的视频文件后，可以向调度设备发送任务分配请求，以获取合适的转码设备来对视频文件进行转码。

步骤 201 中，调度设备在接收到任务分配请求之后，可以获取与调度设备连接的转码设备的负载信息。本发明实施例中，获取转码设备的负载信息的方式有多种，一种可能的实施方式为，调度设备获取所有转码设备的负载信息。例如，若与调度设备连接的转码设备有 10 个，则调度设备可以获取这 10 个转码设备中每个转码设备的负载信息。

另一种可能的实现方式为，调度设备获取特定的转码设备的负载信息。例如，调度设备在接收到任务分配请求后，可以根据任务分配请求中包括的客户端设备的位置信息，从与调度设备连接的所有转码设备中确定出与客户端设备的位置信息相关联的转码设备，进而获取与客户端设备的位置信息相关联的转码设备的负载信息。举个例子，如表 1 所示，为获取特定的转码设备的一种示例。与调度设备连接的转码设备有 10 个，每个转码设备对应的位置信息如表 1

中示出的内容，此处不再一一描述。若客户端设备的位置信息是上海市闵行区的，则可以根据表 1 示出的内容，将这 10 个转码设备中位置信息是上海的转码设备选择出来，即转码设备 3、转码设备 6 和转码设备 9，进而，调度设备可以获取转码设备 3、转码设备 6 和转码设备 9 的负载信息。

表 1：获取特定的转码设备的一种示例

编号	转码设备的位置信息	编号	转码设备的位置信息
转码设备 1	北京市 海淀区	转码设备 6	上海市 普陀区
转码设备 2	杭州市 萧山区	转码设备 7	南京市 浦口区
转码设备 3	上海市 静安区	转码设备 8	广州市 番禺区
转码设备 4	北京市 朝阳区	转码设备 9	上海市 闵行区
转码设备 5	南京市 玄武区	转码设备 10	广州市 白云区

又例如，调度设备在接收到任务分配请求后，可以根据与调度设备连接的转码设备发送的当前状态，从所有转码设备中确定出当前状态为可用状态的转码设备，进而获取可用状态的转码设备的负载信息。或者，也可以根据与调度设备连接的转码设备发送的健康状态，从所有转码设备中确定出为健康状态为健康的转码设备。举个例子，当转码设备的加权任务数大于或等于预设阈值时，认为该转码设备的健康状态为不健康，相应地，可设置该转码设备的健康值 $Q=100$ ；当转码设备的加权任务数小于预设阈值时，认为该转码设备的健康状态为健康，相应地，可设置该转码设备的健康值 $Q=M$ ，其中， M 为该转码设备的平均 CPU 使用率。如此，当一个新的任务接入调度设备后，调度设备可以查询各转码设备的健康状态，并将健康的转码设备的 IP 地址返回给客户端设备，从而使得该任务在健康的转码设备上转码；进一步地，若同时存在多个健康的转码设备，则可以根据健康值 Q 的大小，将健康值较小的转码设备的 IP 地址返回给客户端设备，从而使得该任务在健康的转码设备上转码。

在其它可能的实施例，调度设备在接收到任务分配请求后，也可以通过其它方式来获取多个转码设备的负载信息，具体不做限定。

进一步地，每个转码设备的负载信息可以包括多种信息。例如，负载信息可以包括转码设备承载的转码任务的目标需求信息，而所述转码任务的目标需

求信息又可以包括所述转码任务的目标分辨率和所述转码任务的目标帧率。其中，目标分辨率可以是指转码任务对应的视频（或音频等其它形式）转码后的分辨率，目标帧率可以是指转码任务中画面每秒显示帧数。举个例子，转码任务 A 要求将视频 1 从 640*360 的分辨率转码到 1280*720 的分辨率，且视频 1 的画面每秒显示帧数为 25fps，则对于转码任务 A 而言，目标分辨率为 1280*720，目标帧率为 25fps；再举个例子，转码任务 B 要求将视频 2 从 1280*720 的分辨率转码到 848*480 的分辨率，且视频 2 的画面每秒显示帧数为 29.97fps，则对于转码任务 B 而言，目标分辨率为 848*480，目标帧率为 29.97fps。

或者，负载信息也可以包括每个转码设备中多个执行核心的使用率。执行核心的使用率可以是指每个转码设备的 CPU 中的每个执行核心的使用率。举个例子，转码设备 1 是四核的 CPU，即转码设备 1 具有 4 个执行核心。如表 2 所示，为转码设备 1 中执行核心的信息的一种示例。根据表 2 示出的内容可以得知，转码设备 1 中执行核心 1 的使用率为 40%，转码设备 1 中执行核心 2 的使用率为 5%，转码设备 1 中执行核心 3 的使用率为 30%，转码设备 1 中执行核心 4 的使用率为 10%。

表 2：转码设备 1 中执行核心的信息的一种示例

执行核心的编号	执行核心的使用率
执行核心 1	40%
执行核心 2	5%
执行核心 3	30%
执行核心 4	10%

或者，负载信息还可以包括每个转码设备承载的转码任务的标识。其中，转码任务的标识可以以多种形式来表现，比如，转码任务的标识可以是转码任务的统一资源定位符（Uniform Resource Locator，URL）；又比如，转码任务的标识也可以是转码任务的任务名称，具体不做限定。

本发明实施例中，任务分配请求可以包括待分配任务的目标需求信息和待分配任务的标识。其中，待分配任务的标识可以以 URL 的形式来表现，比如，

待分配任务 1 的标识为“http://a.b.com/shipin/12345.html”，待分配任务的目标需求信息可以包括待分配任务的目标分辨率和待分配任务的目标帧率，比如，待分配任务 1 的目标分辨率为 1280*720，目标帧率为 25fps。

在执行步骤 202 之前，调度设备还可以根据待分配任务的目标分辨率、待分配任务的目标帧率和待分配任务的标识，判断多个转码设备中是否承载有与待分配任务的目标分辨率、待分配任务的目标帧率和待分配任务的标识均一致的转码任务，若存在该转码任务，则调度设备可以将该转码任务对应的转码设备作为目标转码设备，以使客户端设备与目标转码设备建立连接；若不存在该转码任务，则调度设备可以执行上述步骤 202。

如此，调度设备能够优先判断多个转码设备中是否已存在与待分配任务一致的转码任务，从而避免在待分配任务已存在的情况下重复创建转码任务，导致资源浪费的情况。

步骤 202 中，第一转码设备的当前负载值可以根据第一转码设备中的至少一个转码任务的目标需求信息以及所述第一转码设备的设备性能来确定，具体的确定方式有多种。一个示例中，调度设备可以根据第一转码设备的设备性能，确定第一转码设备承载的各转码任务的各目标需求信息对应的预设最大任务数量，进而可以根据第一转码设备中每个目标需求信息对应的预设最大任务数量，以及第一转码设备中每个目标需求信息对应的转码任务的数量，确定第一转码设备的当前负载值。换言之，要获取第一转码设备的当前负载值，需要先获取第一转码设备中不同的目标需求信息能够转码的最大路数（该最大路数也可以看作是一种阈值）。本发明实施例中可以通过压测数据，得到每个目标需求信息能够转码的最大路数，比如，分辨率 F_1 且帧率 Z_1 能转 X_1 路，分辨率 F_2 且帧率 Z_2 能转 X_2 路，其他分辨率同样按照此方法获得能够转码的最大路数，如此，可以确定每个转码设备的当前负载值（该当前负载值事实上可以看作是各个转码设备的加权任务数）。

具体来说，如表 3 所示，为第一转码设备中的转码任务的一种示例。其中，

转码任务 1 对应的目标分辨率为 1280*720, 目标帧率为 25fps; 转码任务 2 对应的目标分辨率为 640*360, 目标帧率为 29.97fps; 转码任务 3 对应的目标分辨率为 848*480, 目标帧率为 25fps; 转码任务 4 对应的目标分辨率为 1280*720, 目标帧率为 25fps; 转码任务 5 对应的目标分辨率为 1280*720, 目标帧率为 29.97fps。根据表 3 示出的内容可知, 第一转码设备中共有 4 种目标需求信息, 分别为: 目标需求信息 1 对应目标分辨率 1280*720 和目标帧率 25fps、目标需求信息 2 对应目标分辨率 640*360 和目标帧率 29.97fps、目标需求信息 3 对应目标分辨率 848*480 和目标帧率 25fps、目标需求信息 4 对应目标分辨率 1280*720 和目标帧率 29.97fps。若根据第一转码设备的设备性能, 可以得到该转码设备中各目标需求信息对应的预设最大任务数量, 具体如表 4 所示。其中, 目标需求信息 1 在第一转码设备中对应的预设最大任务数量为 5 个; 目标需求信息 2 在第一转码设备中对应的预设最大任务数量为 15 个; 目标需求信息 3 在第一转码设备中对应的预设最大任务数量为 10 个; 目标需求信息 4 在第一转码设备中对应的预设最大任务数量为 5 个。如此, 可以根据表 3 中示出的每个目标需求信息对应的任务数量, 和表 4 中示出的每个目标需求信息对应的预设最大任务数量, 采用公式(1)可以计算得到该转码设备的当前负载值为 0.77。

$$A=f_1 \times \frac{1}{X_1} + f_2 \times \frac{1}{X_2} + \dots + f_n \times \frac{1}{X_n} \quad \text{公式 (1)}$$

其中, A 为第一转码设备的当前负载值; f_i 为第一转码设备中第 i 个目标需求信息对应的任务数量, X_i 为第一转码设备中第 i 个目标需求信息对应的预设最大任务数量, $i=1, 2, \dots, n$; n 为第一转码设备承载的转码任务的目标需求信息的数目, n 为大于或等于 1 的整数。

需要说明的是, 公式 (1) 仅为一种示例, 本领域技术人员也可以通过其它方式来确定当前负载值, 比如在公式 (1) 的基础上考虑各目标需求信息的权值, 具体不做限定。

如此, 调度设备可以根据各目标需求信息在第一转码设备中的任务数量以及预设最大任务数量, 得到第一转码设备的当前负载值, 这种方法能够更加合

理地计算转码设备的当前负载值，从而提高确定转码设备的合理性。

表 3：第一转码设备中的转码任务的一种示例

转码任务的编号	目标分辨率	目标帧率
转码任务 1	1280*720	25fps
转码任务 2	640*360	29.97fps
转码任务 3	848*480	25fps
转码任务 4	1280*720	25fps
转码任务 5	1280*720	29.97fps

表 4：第一转码设备中各目标需求信息的预设最大任务数量的一种示例

目标需求信息	预设最大任务数量
目标需求信息 1	5
目标需求信息 2	15
目标需求信息 3	10
目标需求信息 4	5

另一个示例中，调度设备可以根据第一转码设备的设备性能以及第一转码设备中每个转码任务的目标需求信息，确定每个目标需求信息在第一转码设备中的性能占用率，并根据每个目标需求信息对应的转码任务的数量，确定每个目标需求信息在第一转码设备中的总性能占用率，进而得到第一转码设备的当前负载值。

具体来说，以表 3 示出的第一转码设备中的转码任务为例，若第一转码设备中目标需求信息 1（即目标分辨率 1280*720 和目标帧率 25fps）在该转码设备中性能占用率为 20%，目标需求信息 2（即目标分辨率 640*360 和目标帧率 29.97fps）在第一转码设备中性能占用率为 5%，目标需求信息 3（即目标分辨率 848*480 和目标帧率 25fps）在第一转码设备中性能占用率为 10%，目标需求信息 4（即目标分辨率 1280*720 和目标帧率 29.97fps）在第一转码设备中性能占用率为 25%，则根据表 3 中示出的每个目标需求信息对应的转码任务的数量，采用公式（2）可以计算得到第一转码设备的当前负载值为 76%。

$$A=f_1 \times C_1 + f_2 \times C_2 + \dots + f_n \times C_n \quad \text{公式（2）}$$

其中，A 为第一转码设备的当前负载值； f_i 为第一转码设备中第 i 个目标

需求信息对应的任务数量, C_i 为第一转码设备中第 i 个目标需求信息在该转码设备中性能占用率, $i=1, 2, \dots, n$; n 为第一转码设备承载的转码任务的目标需求信息的数目, n 为大于或等于 1 的整数。

需要说明的是, 公式 (2) 仅为一种示例, 本领域技术人员也可以通过其它方式来确定当前负载值, 比如在公式 (2) 的基础上考虑各目标需求信息的权值, 具体不做限定。

在其它示例中, 本领域技术人员也可以根据经验和实际情况来确定第一转码设备的当前负载值, 具体不做限定。

步骤 203 中, 调度设备确定目标转码设备的方式有多种。一种可能的实现方式为, 调度设备可以从多个转码设备中确定出当前负载值小于第一阈值的候选转码设备, 若只存在一个候选转码设备, 则可以将该候选转码设备作为目标转码设备; 若存在多个候选转码设备, 则可以将当前负载值最小的候选转码设备作为目标转码设备, 或者也可以将多个候选转码设备中的任一转码设备作为目标转码设备, 具体不做限制。

需要说明的是, 本领域技术人员可以根据经验和实际情况来设置第一阈值, 具体不做限定。

另一种可能的实现方式为, 调度设备可以根据多个转码设备的当前负载值和多个转码设备的平均 CPU 使用率, 从多个转码设备中确定出当前负载值小于第一阈值, 且平均 CPU 使用率小于第二阈值的目标转码设备, 其中, 平均 CPU 使用率可以是根据负载信息中包括的每个转码设备中多个执行核心的使用率确定的。具体来说, 调度设备可以先从多个转码设备中确定出当前负载值小于第一阈值的至少一个候选转码设备, 然后再从至少一个候选转码设备中确定出平均 CPU 使用率小于第二阈值的目标转码设备; 或者, 调度设备也可以先从多个转码设备中确定出平均 CPU 使用率小于第二阈值的至少一个候选转码设备, 然后再从至少一个候选转码设备中确定出当前负载值小于第一阈值的目标转码设备。

需要说明的是，本领域技术人员可以根据经验和实际情况来设置第二阈值，具体不做限定。

如此，在确定目标转码设备时，既考虑每个转码设备的当前负载值，又考虑每个转码设备的平均CPU使用率，从而能够从不同的角度来选择转码设备，提高确定转码设备的合理性。

进一步地，确定平均 CPU 使用率的方式可以有多种。举个例子，可以将第一转码设备中的多个执行核心按照每个执行核心的使用率从大到小进行排名，然后可以将排名前 N 位的执行核心的使用率的平均值确定为第一转码设备的平均 CPU 使用率，N 为大于或等于 1 的整数。以表 2 中示出的转码设备 1 中执行核心的信息为例，可以将转码设备 1 中的 4 个执行核心按照每个执行核心的使用率从大到小进行排名：执行核心 1>执行核心 3>执行核心 4>执行核心 2；若设定 N=2，则根据表 2 中示出的每个执行核心的使用率，采用公式(3)可以计算得到转码设备 1 的平均 CPU 使用率为 35%。

$$M = \frac{top_1 + top_2 + \dots + top_N}{N} \quad \text{公式 (3)}$$

其中，M 为第一转码设备的平均 CPU 使用率； top_j 为排名第 j 位的执行核心的使用率， $j=1, 2, \dots, N$ ；N 为大于或等于 1 的整数。

如此，能够避免转码设备中执行核心过多时计算量大的问题，从而降低平均CPU使用率的计算量，进而降低调度设备的运行负担。

需要说明的是，上述所描述的方法仅为一种示例，本领域技术人员也可以根据经验和实际情况采用其它方式来确定平均 CPU 使用率。例如，可以对当前时刻下的 m 核 CPU 中每个执行核心的使用率从大到小进行排列，然后取使用率最大的 $(1/4 * m)$ 个核的均值作为第一转码设备的平均 CPU 使用率，相应地，上文所示的公式 (3) 也可以修改为公式 (4)，具体不再赘述。

$$M = \frac{top_1 + top_2 + \dots + top_{m/4}}{m/4} \quad \text{公式 (4)}$$

其中，M 为第一转码设备的平均 CPU 使用率；m 为第一转码设备中执行

核心的个数； top_j 为排名第 j 位的执行核心的使用率， $j=1, 2, \dots, m/4$ ； m 为大于或等于1的整数。

再举个例子，也可以将第一转码设备中的多个执行核心的平均值作为第一转码设备的平均CPU使用率。以表2中示出的转码设备1中执行核心的信息为例，可以将转码设备1中的4个执行核心的使用率的平均值作为转码设备1的平均CPU使用率，即转码设备1的平均CPU使用率为42.5%。这种方法计算较为便捷。

另一种可能的实现方式为，调度设备可以根据多个转码设备的当前负载值和多个转码设备的预测负载值，从多个转码设备中确定出当前负载值小于第一阈值，且预测负载值小于第三阈值的目标转码设备，其中，预设负载值可以根据转码设备的当前负载值和待分配转码任务的负载值确定的，而待分配任务的负载值又可以根据所述待分配转码任务的目标需求信息确定的。具体来说，调度设备可以先从多个转码设备中确定出当前负载值小于第一阈值的至少一个候选转码设备，然后再从至少一个候选转码设备中确定出预设负载值小于第三阈值的目标转码设备；或者，调度设备也可以先从多个转码设备中确定出预设负载值小于第三阈值的至少一个候选转码设备，然后再从至少一个候选转码设备中确定出当前负载值小于第一阈值的目标转码设备。

需要说明的是，本领域技术人员可以根据经验和实际情况来设置第三阈值，具体不做限定。

如此，在确定目标转码设备时，既考虑每个转码设备的当前负载值，又考虑每个转码设备在接收待分配转码任务后的预测负载值，从而能够从不同的角度来选择转码设备，提高确定转码设备的合理性。

进一步地，确定预测负载值的方式可以有多种。一个示例中，调度设备可以根据待分配任务的目标需求信息在多个转码设备中分别对应的预设最大任务数量，确定待分配任务在多个转码设备中分别对应的负载值，再结合每个转码设备的当前负载值，可以确定每个转码设备的预测负载值。举个例子，如表

5 所示，为多个转码设备的负载信息的一种示例。其中，转码设备 1 的当前负载值为 0.77，待分配任务的目标需求信息在转码设备 1 中对应的预设最大任务数量为 3 个，则待分配任务在转码设备 1 中对应的负载值可以采用公式（5）计算，得到待分配任务在转码设备 1 中对应的负载值为 0.33，进而可以根据公式（6）确定转码设备 1 的预测负载值为 1.1；其余转码设备的预测负载值的计算方法可以参考转码设备 1，在此不再详细描述。

$$P = \frac{1}{X} \quad \text{公式 (5)}$$

其中，P 为待分配任务在第一转码设备中对应的负载值；X 为待分配任务的目标需求信息在第一转码设备中对应的预设最大任务数量。

$$Q = P + A \quad \text{公式 (6)}$$

其中，Q 为第一转码设备的预测负载值；P 为待分配任务在第一转码设备中对应的负载值；A 为第一转码设备的当前负载值。

表 5：多个转码设备的负载信息的一种示例

编号	当前负载值	预设最大任务数量	预测负载值
转码设备 1	0.77	3	1.1
转码设备 2	0.2	5	0.4
转码设备 3	0.55	10	0.65
转码设备 4	0.9	5	1.1
转码设备 5	0.1	3	0.43

另一个示例中，调度设备可以将待分配任务的目标需求信息在多个转码设备中分别对应的性能占有率作为待分配任务在多个转码设备中分别对应的负载值，再结合每个转码设备的当前负载值，可以确定每个转码设备的预测负载值。

另一种可能的实现方式为，调度设备可以根据多个转码设备的当前负载值、多个转码设备的平均 CPU 使用率和多个转码设备的预测负载值，从多个转码设备中确定出当前负载值小于第一阈值，且平均 CPU 使用率小于第二阈值，且预测负载值小于第三阈值的目标转码设备。具体来说，调度设备可以先从多个转码设备中确定出当前负载值小于第一阈值的至少一个初始候选转码设备，

然后从至少一个初始候选转码设备中确定出平均 CPU 使用率小于第二阈值的至少一个候选转码设备，再从至少一个候选转码设备中确定出预设负载值小于第三阈值的目标转码设备；或者，调度设备可以先从多个转码设备中确定出当前负载值小于第一阈值的至少一个初始候选转码设备，然后从至少一个初始候选转码设备中确定出预设负载值小于第三阈值的至少一个候选转码设备，再从至少一个候选转码设备中确定出平均 CPU 使用率小于第二阈值的目标转码设备；或者，调度设备可以先从多个转码设备中确定出平均 CPU 使用率小于第二阈值的至少一个初始候选转码设备，然后从至少一个初始候选转码设备中确定出当前负载值小于第一阈值的至少一个候选转码设备，再从至少一个候选转码设备中确定出预设负载值小于第三阈值的目标转码设备，具体不做限定。

如此，在确定目标转码设备时，既可以考虑每个转码设备的当前负载值，又可以考虑每个转码设备的平均CPU使用率，还可以考虑每个转码设备在接收待转码任务后的预测负载值，从而能够从不同的角度来选择转码设备，提高确定转码设备的合理性。

实施例二：

基于图 1 所示的系统架构，图 3 示例性示出了本发明实施例提供的另一种转码任务的分配方法所对应的流程示意图，包括以下步骤：

步骤301，调度设备向转码设备发送状态请求。

步骤302，转码设备接收到调度设备发送的状态请求后，根据自身的负载信息以及自身的设备性能，确定当前负载值。

步骤303，转码设备根据当前负载值，生成状态请求的响应消息。

步骤304，转码设备向调度设备发送的所述响应消息。

步骤305，调度设备根据所述响应消息确定出目标转码设备，所述目标转码设备用于执行待分配的转码任务。

如此，转码设备可以根据自身的负载信息来计算当前负载值，可以减轻调度设备的计算量，从而减轻调度设备的运行负担；进一步地，调度设备可以选

择当前负载值较小的转码设备作为目标转码设备，相比于现有技术中直接根据转码设备的IP地址来选择目标转码设备的方式而言，本发明实施例充分考虑了转码设备的负载情况，有效避免了因转码设备过载而不能及时执行转码任务的情况，提高了视频转码的效率。

步骤301和步骤302中，状态请求可以包括待分配任务的目标需求信息和待分配任务的标识。其中，待分配任务的标识可以以URL的形式来表现，比如，待分配任务1的标识为“<http://a.b.com/shipin/12345.html>”，待分配任务的目标需求信息可以包括待分配任务的目标分辨率和待分配任务的目标帧率，比如，待分配任务1的目标分辨率为1280*720，目标帧率为25fps。

进一步地，转码设备在接收到调度设备发送的状态请求之后，可以先根据待分配任务的目标分辨率、待分配任务的目标帧率和待分配任务的标识，判断转码设备承载的转码任务中是否存在与待分配任务的目标分辨率、待分配任务的目标帧率和待分配任务的标识均一致的转码任务，若存在该转码任务，则可以向调度设备发送指示信息，以使调度设备将该转码设备作为目标转码设备；若不存在该转码任务，则转码设备可以继续执行上述步骤302。

如此，转码设备能够优先判断是否已存在与待分配任务一致的转码任务，从而避免在待分配任务已存在的情况下重复创建转码任务，导致资源浪费的情况。

本发明实施例中，负载信息可以包括多种信息。比如，负载信息可以包括转码设备承载的转码任务的目标需求信息，而所述转码任务的目标需求信息又可以包括所述转码任务的目标分辨率和所述转码任务的目标帧率。其中，目标分辨率可以是指转码任务对应的视频（或音频等其它形式）转码后的分辨率，目标帧率可以是指转码任务中画面每秒显示帧数。

又比如，负载信息还可以包括每个转码设备中多个执行核心的使用率，执行核心的使用率可以是指每个转码设备的CPU中的每个执行核心的使用率。

又比如，负载信息还可以包括承载的转码任务的标识。其中，转码任务的

标识可以是转码任务的 URL；或者，也可以是转码任务的任务名称，具体不做限定。

需要说明的是，关于负载信息的具体示例可以参考上述实施例一种示出的内容，此处不再详细描述。

下面具体介绍转码设备根据自身的负载信息进行计算的方式。

一个示例中，转码设备可以根据负载信息中各转码任务的目标需求信息以及自身的设备性能，确定当前负载值。举个例子，转码设备可以根据设备性能，确定所承载的各转码任务的各目标需求信息对应的预设最大任务数量，然后再根据每个目标需求信息对应的预设最大任务数量，以及每个目标需求信息对应的转码任务的数量，确定转码设备的当前负载值。具体地，当前负载值可以通过实施例一中的公式（1）得到，此处不再详细描述。需要说明的是，公式（1）仅为一种示例，本领域技术人员也可以通过其它方式来确定当前负载值，比如在公式（1）的基础上考虑各目标需求信息的权值，具体不做限定。采用这种方法，能够更加合理地计算转码设备的当前负载值，从而提高确定转码设备的合理性。

再举个例子，转码设备可以根据设备性能以及每个转码任务的目标需求信息，确定每个目标需求信息在所述转码设备中的性能占用率，并根据每个目标需求信息对应的转码任务的数量，确定每个目标需求信息在所述转码设备中的总性能占用率，进而得到转码设备的当前负载值。具体地，当前负载值可以通过实施例一中的公式（2）得到，此处不再详细描述。需要说明的是，公式（2）仅为一种示例，本领域技术人员也可以通过其它方式来确定当前负载值，比如在公式（2）的基础上考虑各目标需求信息的权值，具体不做限定。

又一个示例中，转码设备可以根据负载信息中多个执行核心的使用率以及自身的设备性能，确定平均 CPU 使用率。

进一步地，确定平均 CPU 使用率的方式可以有多种。举个例子，转码设备可以将多个执行核心按照每个执行核心的使用率从大到小进行排名，然后可

以将排名前 N 位的执行核心的使用率的平均值确定为平均 CPU 使用率， N 为大于或等于 1 的整数。具体地，平均 CPU 使用率可以通过实施例一中的公式 (3) 得到，此处不再详细描述。采用这种方法，能够避免转码设备中执行核心过多时计算量大的问题，从而降低平均 CPU 使用率的计算量，进而降低转码设备的运行负担。

需要说明的是，上述所描述的方法仅为一种示例，本领域技术人员也可以根据经验和实际情况采用其它方式来确定平均 CPU 使用率。例如，可以对当前时刻下的 m 核 CPU 中每个执行核心的使用率从大到小进行排列，然后取使用率最大的 $(1/4 * m)$ 个核的均值作为第一转码设备的平均 CPU 使用率，具体可以参考实施例一中的公式 (4) 此处不再详细描述。

再举个例子，转码设备也可以将多个执行核心的平均值作为平均 CPU 使用率。这种方法计算较为便捷。

又一个示例中，转码设备可以根据待分配任务的目标需求信息和所述转码设备的当前负载值，确定自身的预测负载值。

进一步地，确定预测负载值的方式可以有多种。举个例子，转码设备可以根据待分配任务的目标需求信息在该转码设备中对应的预设最大任务数量，确定待分配任务的负载值，再结合该转码设备的当前负载值，可以确定该转码设备的预测负载值。具体地，预测负载值可以通过实施例一中的公式 (5) 和公式 (6) 得到，此处不再详细描述。

再举个例子，转码设备可以将待分配任务的目标需求信息在该转码设备中分别对应的性能占有率作为待分配任务的负载值，再结合该转码设备的当前负载值，可以确定该转码设备的预测负载值。

步骤303和步骤304中，转码设备生成的状态请求的响应消息可以包括多种形式的内容。比如，状态请求的响应消息可以仅包括转码设备的当前负载值；或者，状态请求的响应消息可以包括转码设备的当前负载值和平均CPU使用率；或者，状态请求的响应消息可以包括转码设备的当前负载值和预测负载值；或

者，状态请求的响应消息可以包括转码设备的当前负载值、平均CPU使用率和预测负载值，具体不做限定。

相应地，步骤305中，调度设备接收到的状态请求的响应消息中包括的内容的不同，确定目标转码设备的方式也不同。

一个示例中，响应消息包括转码设备的当前负载值。调度设备可以从多个转码设备中确定出当前负载值小于第一阈值的候选转码设备，若只存在一个候选转码设备，则可以将该候选转码设备作为目标转码设备；若存在多个候选转码设备，则可以将当前负载值最小的候选转码设备作为目标转码设备，或者也可以将多个候选转码设备中的任一转码设备作为目标转码设备，具体不做限制。

需要说明的是，本领域技术人员可以根据经验和实际情况来设置第一阈值，具体不做限定。

另一个示例中，响应消息包括转码设备的当前负载值和平均CPU使用率。调度设备可以根据多个转码设备的当前负载值和多个转码设备的平均CPU使用率，从多个转码设备中确定出当前负载值小于第一阈值，且平均CPU使用率小于第二阈值的目标转码设备。具体来说，调度设备可以先从多个转码设备中确定出当前负载值小于第一阈值的至少一个候选转码设备，然后再从至少一个候选转码设备中确定出平均CPU使用率小于第二阈值的目标转码设备；或者，调度设备也可以先从多个转码设备中确定出平均CPU使用率小于第二阈值的至少一个候选转码设备，然后再从至少一个候选转码设备中确定出当前负载值小于第一阈值的目标转码设备。

需要说明的是，本领域技术人员可以根据经验和实际情况来设置第二阈值，具体不做限定。

如此，在确定目标转码设备时，既考虑每个转码设备的当前负载值，又考虑每个转码设备的平均CPU使用率，从而能够从不同的角度来选择转码设备，提高确定转码设备的合理性。

另一个示例中，响应消息包括转码设备的当前负载值和预测负载值。调度

设备可以根据多个转码设备的当前负载值和多个转码设备的预测负载值，从多个转码设备中确定出当前负载值小于第一阈值，且预测负载值小于第三阈值的目标转码设备。具体来说，调度设备可以先从多个转码设备中确定出当前负载值小于第一阈值的至少一个候选转码设备，然后再从至少一个候选转码设备中确定出预设负载值小于第三阈值的目标转码设备；或者，调度设备也可以先从多个转码设备中确定出预设负载值小于第三阈值的至少一个候选转码设备，然后再从至少一个候选转码设备中确定出当前负载值小于第一阈值的目标转码设备。

需要说明的是，本领域技术人员可以根据经验和实际情况来设置第三阈值，具体不做限定。

如此，在确定目标转码设备时，既考虑每个转码设备的当前负载值，又考虑每个转码设备在接收待分配转码任务后的预测负载值，从而能够从不同的角度来选择转码设备，提高确定转码设备的合理性。

另一个示例中，响应消息包括转码设备的当前负载值、平均 CPU 使用率和预测负载值。调度设备可以根据多个转码设备的当前负载值、多个转码设备的平均 CPU 使用率和多个转码设备的预测负载值，从多个转码设备中确定出当前负载值小于第一阈值，且平均 CPU 使用率小于第二阈值，且预测负载值小于第三阈值的目标转码设备。具体来说，调度设备可以先从多个转码设备中确定出当前负载值小于第一阈值的至少一个初始候选转码设备，然后从至少一个初始候选转码设备中确定出平均 CPU 使用率小于第二阈值的至少一个候选转码设备，再从至少一个候选转码设备中确定出预设负载值小于第三阈值的目标转码设备；或者，调度设备可以先从多个转码设备中确定出当前负载值小于第一阈值的至少一个初始候选转码设备，然后从至少一个初始候选转码设备中确定出预设负载值小于第三阈值的至少一个候选转码设备，再从至少一个候选转码设备中确定出平均 CPU 使用率小于第二阈值的目标转码设备；或者，调度设备可以先从多个转码设备中确定出平均 CPU 使用率小于第二阈值的至少

一个初始候选转码设备，然后从至少一个初始候选转码设备中确定出当前负载值小于第一阈值的至少一个候选转码设备，再从至少一个候选转码设备中确定出预设负载值小于第三阈值的目标转码设备，具体不做限定。

如此，在确定目标转码设备时，既可以考虑每个转码设备的当前负载值，又可以考虑每个转码设备的平均CPU使用率，还可以考虑每个转码设备在接收待转码任务后的预测负载值，从而能够从不同的角度来选择转码设备，提高确定转码设备的合理性。

实施例三：基于同样的发明构思，图4示例性示出了本发明实施例提供一种调度设备的结构示意图，如图4所示，所述调度设备400包括接收单元401、处理单元402和发送单元403。

接收单元401，用于接收到客户端设备发送的任务分配请求后，获取多个转码设备的负载信息，每个转码设备的负载信息包括转码设备承载的转码任务的目标需求信息，所述转码任务的目标需求信息包括所述转码任务的目标分辨率和所述转码任务的目标帧率；

处理单元402，用于针对第一转码设备，根据所述第一转码设备中的至少一个转码任务的目标需求信息以及所述第一转码设备的设备性能，确定所述第一转码设备的当前负载值，所述第一转码设备为所述多个转码设备中的任一设备；以及根据所述多个转码设备的当前负载值，从所述多个转码设备中确定出当前负载值小于第一阈值的目标转码设备；

发送单元403，用于向所述客户端设备返回所述任务分配请求的响应消息，所述响应消息用于指示所述客户端设备向所述目标转码设备发送转码任务。

在一种可能的实现方式中，所述处理单元402具体用于：

根据所述第一转码设备的设备性能，确定所述第一转码设备承载的各转码任务的各目标需求信息对应的预设最大任务数量；

根据所述第一转码设备中每个目标需求信息对应的预设最大任务数量，以及所述第一转码设备中每个目标需求信息对应的转码任务的数量，确定所述第

一转码设备的当前负载值。

在一种可能的实现方式中，所述当前负载值通过以下方式确定：

$$A=f_1\times\frac{1}{X_1}+f_2\times\frac{1}{X_2}+\cdots+f_n\times\frac{1}{X_n}$$

其中，A 为所述第一转码设备的当前负载值； f_i 为所述第一转码设备中第 i 个目标需求信息对应的任务数量， X_i 为所述第一转码设备中第 i 个目标需求信息对应的预设最大任务数量， $i=1, 2, \dots, n$ ；n 为所述第一转码设备承载的转码任务的目标需求信息的数目，n 为大于或等于 1 的整数。

在一种可能的实现方式中，所述每个转码设备的负载信息还包括所述每个转码设备中多个执行核心的使用率；

所述处理单元402具体用于：

根据所述多个转码设备的当前负载值和所述多个转码设备的平均中央处理器CPU使用率，从所述多个转码设备中确定出当前负载值小于第一阈值，且平均CPU使用率小于第二阈值的目标转码设备；所述转码设备的平均CPU使用率是根据转码设备中的多个执行核心的使用率确定的。

在一种可能的实现方式中，所述处理单元402具体用于：

将所述第一转码设备中的多个执行核心按照每个执行核心的使用率从大到小进行排名；

将排名前N位的执行核心的使用率的平均值确定为所述第一转码设备的平均CPU使用率，N为大于或等于1的整数。

在一种可能的实现方式中，所述处理单元402具体用于：

从所述多个转码设备中确定出当前负载值小于所述第一阈值的至少一个候选转码设备；

根据每个候选转码设备中的多个执行核心的使用率，确定所述每个候选转码设备的平均CPU使用率；

从所述至少一个候选转码设备中确定出平均CPU使用率小于所述第二阈值的的目标转码设备。

在一种可能的实现方式中，所述任务分配请求包括待分配任务的目标需求信息，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率和所述待分配任务的目标帧率；

所述处理单元402具体用于：

根据所述多个转码设备的当前负载值和所述多个转码设备的预测负载值，从所述多个转码设备中确定出当前负载值小于第一阈值，且预测负载值小于第三阈值的目标转码设备；所述转码设备的预设负载值是根据所述转码设备的当前负载值和所述待分配转码任务的负载值确定的，所述待分配任务的负载值是根据所述待分配转码任务的目标需求信息确定的。

在一种可能的实现方式中，所述任务分配请求包括待分配任务的目标需求信息和所述待分配任务的标识，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率、所述待分配任务的目标帧率；所述每个转码设备的负载信息还包括转码设备承载的转码任务的标识；

所述处理单元402在确定所述第一转码设备的当前负载值之前，还用于：

确定所述多个转码设备承载的转码任务中不存在与所述待分配任务的目标分辨率、所述待分配任务的目标帧率和所述待分配任务的标识均一致的转码任务。

实施例四：

基于同样的发明构思，图5示例性示出了本发明实施例提供一种转码设备的结构示意图，如图5所示，所述转码设备500包括接收单元501、处理单元502和发送单元503。

接收单元501，用于接收到调度设备发送的状态请求；

处理单元502，用于根据自身的负载信息以及自身的设备性能，确定当前负载值；所述负载信息包括所述转码设备承载的至少一个转码任务的目标需求信息，所述转码任务的目标需求信息包括所述转码任务的目标分辨率和所述转码任务的目标帧率；以及根据所述当前负载值，生成所述状态请求的响应消息；

发送单元503，用于向所述调度设备发送的所述响应消息，以使所述调度设备根据所述响应消息确定出目标转码设备，所述目标转码设备用于执行待分配的转码任务。

在一种可能的实现方式中，所述处理单元502具体用于：

根据所述设备性能，确定所承载的各转码任务的各目标需求信息对应的预设最大任务数量；

根据每个目标需求信息对应的预设最大任务数量，以及所述每个目标需求信息对应的转码任务的数量，确定所述转码设备的当前负载值。

在一种可能的实现方式中，所述当前负载值通过以下方式确定：

$$A=f_1\times\frac{1}{X_1}+f_2\times\frac{1}{X_2}+\cdots+f_n\times\frac{1}{X_n}$$

其中，A 为所述转码设备的当前负载值； f_i 为所述转码设备中第 i 个目标需求信息对应的任务数量， X_i 为所述转码设备中第 i 个目标需求信息对应的预设最大任务数量， $i=1, 2, \dots, n$ ；n 为所述转码设备承载的转码任务的目标需求信息的数目，n 为大于或等于 1 的整数。

在一种可能的实现方式中，所述负载信息还包括多个执行核心的使用率；所述处理单元502还用于：

根据所述多个执行核心的使用率，确定平均CPU使用率；

根据所述当前负载值和所述确定平均CPU使用率，生成所述状态请求的响应消息。

在一种可能的实现方式中，所述处理单元502具体用于：

将所述多个执行核心按照每个执行核心的使用率从大到小进行排名；

将排名前N位的执行核心的使用率的平均值确定为所述平均CPU使用率，N为大于或等于1的整数。

在一种可能的实现方式中，所述状态请求包括所述待分配任务的目标需求信息，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率和所述待分配任务的目标帧率；

所述处理单元502还用于：

根据所述待分配任务的目标需求信息，确定所述待分配任务的负载值；

根据所述当前负载值和所述待分配任务的负载值，确定自身的预测负载值；

根据所述当前负载值和所述预测负载值，生成所述状态请求的响应消息。

在一种可能的实现方式中，所述状态请求包括所述待分配任务的目标需求信息和所述待分配任务的标识，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率、所述待分配任务的目标帧率；

所述处理单元在确定所述当前负载值之前，所述处理单元502还用于：

确定所述转码设备承载的转码任务中不存在与所述待分配任务的目标分辨率、所述待分配任务的目标帧率和所述待分配任务的标识均一致的转码任务。

如此，调度设备可以根据每个转码设备的负载信息来计算每个转码设备的负载值，从而可以选择负载值较小的转码设备作为目标转码设备，相比于现有技术中直接根据转码设备的 IP 地址来选择目标转码设备的方式而言，本发明实施例充分考虑了转码设备的负载情况，有效避免了因转码设备过载而不能及时执行转码任务的情况，提高了视频转码的效率。

本领域内的技术人员应明白，本发明的实施例可提供为方法、系统、或计算机程序产品。因此，本发明可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

本发明是参照根据本发明实施例的方法、设备（系统）、和计算机程序产品的流程图和 / 或方框图来描述的。应理解可由计算机程序指令实现流程图和 / 或方框图中的每一流程和 / 或方框、以及流程图和 / 或方框图中的流程和 / 或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一

个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的步骤。

尽管已描述了本发明的优选实施例，但本领域内的技术人员一旦得知了基本创造性概念，则可对这些实施例作出另外的变更和修改。所以，所附权利要求意欲解释为包括优选实施例以及落入本发明范围的所有变更和修改。

显然，本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样，倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内，则本发明也意图包含这些改动和变型在内。

权 利 要 求 书

1、一种转码任务的分配方法，其特征在于，所述方法包括：

调度设备接收到客户端设备发送的任务分配请求后，获取多个转码设备的负载信息，每个转码设备的负载信息包括转码设备承载的转码任务的目标需求信息，所述转码任务的目标需求信息包括所述转码任务的目标分辨率和所述转码任务的目标帧率；

针对第一转码设备，所述调度设备根据所述第一转码设备中的至少一个转码任务的目标需求信息以及所述第一转码设备的设备性能，确定所述第一转码设备的当前负载值，所述第一转码设备为所述多个转码设备中的任一设备；

所述调度设备根据所述多个转码设备的当前负载值，从所述多个转码设备中确定出当前负载值小于第一阈值的目标转码设备，并向所述客户端设备返回所述任务分配请求的响应消息，所述响应消息用于指示所述客户端设备向所述目标转码设备发送转码任务。

2、根据权利要求1所述的方法，其特征在于，所述调度设备根据所述第一转码设备中的至少一个转码任务的目标需求信息以及所述第一转码设备的设备性能，确定所述第一转码设备的当前负载值，包括：

所述调度设备根据所述第一转码设备的设备性能，确定所述第一转码设备承载的各转码任务的各目标需求信息对应的预设最大任务数量；

所述调度设备根据所述第一转码设备中每个目标需求信息对应的预设最大任务数量，以及所述第一转码设备中每个目标需求信息对应的转码任务的数量，确定所述第一转码设备的当前负载值。

3、根据权利要求2所述的方法，其特征在于，所述当前负载值通过以下方式确定：

$$A=f_1\times\frac{1}{X_1}+f_2\times\frac{1}{X_2}+\cdots+f_n\times\frac{1}{X_n}$$

其中，A 为所述第一转码设备的当前负载值； f_i 为所述第一转码设备中第

i 个目标需求信息对应的任务数量, X_i 为所述第一转码设备中第 i 个目标需求信息对应的预设最大任务数量, $i=1, 2, \dots, n$; n 为所述第一转码设备承载的转码任务的目标需求信息的数目, n 为大于或等于 1 的整数。

4、根据权利要求1所述的方法, 其特征在于, 所述每个转码设备的负载信息还包括所述每个转码设备中多个执行核心的使用率;

所述调度设备根据所述多个转码设备的当前负载值, 从所述多个转码设备中确定出当前负载值小于第一阈值的目标转码设备, 包括:

所述调度设备根据所述多个转码设备的当前负载值和所述多个转码设备的平均中央处理器CPU使用率, 从所述多个转码设备中确定出当前负载值小于第一阈值, 且平均CPU使用率小于第二阈值的目标转码设备; 所述转码设备的平均CPU使用率是根据转码设备中的多个执行核心的使用率确定的。

5、根据权利要求4所述的方法, 其特征在于, 所述调度设备根据所述第一转码设备中的多个执行核心的使用率, 确定所述第一转码设备的平均CPU使用率, 包括:

所述调度设备将所述第一转码设备中的多个执行核心按照每个执行核心的使用率从大到小进行排名;

所述调度设备将排名前 N 位的执行核心的使用率的平均值确定为所述第一转码设备的平均CPU使用率, N 为大于或等于1的整数。

6、根据权利要求4所述的方法, 其特征在于, 所述从所述多个转码设备中确定出当前负载值小于第一阈值, 且平均CPU使用率小于第二阈值的目标转码设备, 包括:

所述调度设备从所述多个转码设备中确定出当前负载值小于所述第一阈值的至少一个候选转码设备;

所述调度设备根据每个候选转码设备中的多个执行核心的使用率, 确定所述每个候选转码设备的平均CPU使用率;

所述调度设备从所述至少一个候选转码设备中确定出平均CPU使用率小

于所述第二阈值的目标转码设备。

7、根据权利要求1所述的方法，其特征在于，所述任务分配请求包括待分配任务的目标需求信息，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率和所述待分配任务的目标帧率；

所述调度设备根据所述多个转码设备的当前负载值，从所述多个转码设备中确定出当前负载值小于第一阈值的目标转码设备，包括：

所述调度设备根据所述多个转码设备的当前负载值和所述多个转码设备的预测负载值，从所述多个转码设备中确定出当前负载值小于第一阈值，且预测负载值小于第三阈值的目标转码设备；所述转码设备的预设负载值是根据所述转码设备的当前负载值和所述待分配转码任务的负载值确定的，所述待分配任务的负载值是根据所述待分配转码任务的目标需求信息确定的。

8、根据权利要求1至7中任一项所述的方法，其特征在于，所述任务分配请求包括待分配任务的目标需求信息和所述待分配任务的标识，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率、所述待分配任务的目标帧率；所述每个转码设备的负载信息还包括转码设备承载的转码任务的标识；

在所述获取多个转码设备的负载信息之后，在所述调度设备确定所述第一转码设备的当前负载值之前，所述方法还包括：

所述调度设备确定所述多个转码设备承载的转码任务中不存在与所述待分配任务的目标分辨率、所述待分配任务的目标帧率和所述待分配任务的标识均一致的转码任务。

9、一种转码任务的分配方法，其特征在于，所述方法包括：

转码设备接收到调度设备发送的状态请求后，根据自身的负载信息以及自身的设备性能，确定当前负载值；所述负载信息包括所述转码设备承载的至少一个转码任务的目标需求信息，所述转码任务的目标需求信息包括所述转码任务的目标分辨率和所述转码任务的目标帧率；

所述转码设备根据所述当前负载值，生成所述状态请求的响应消息；

所述转码设备向所述调度设备发送的所述响应消息，以使所述调度设备根据所述响应消息确定出目标转码设备，所述目标转码设备用于执行待分配的转码任务。

10、根据权利要求9所述的方法，其特征在于，所述转码设备根据自身的负载信息以及自身的设备性能，确定当前负载值，包括：

所述转码设备根据所述设备性能，确定所承载的各转码任务的各目标需求信息对应的预设最大任务数量；

所述转码设备根据每个目标需求信息对应的预设最大任务数量，以及所述每个目标需求信息对应的转码任务的数量，确定所述转码设备的当前负载值。

11、根据权利要求10所述的方法，其特征在于，所述当前负载值通过以下方式确定：

$$A=f_1\times\frac{1}{X_1}+f_2\times\frac{1}{X_2}+\cdots+f_n\times\frac{1}{X_n}$$

其中，A 为所述转码设备的当前负载值； f_i 为所述转码设备中第 i 个目标需求信息对应的任务数量， X_i 为所述转码设备中第 i 个目标需求信息对应的预设最大任务数量， $i=1, 2, \dots, n$ ；n 为所述转码设备承载的转码任务的目标需求信息的数目，n 为大于或等于 1 的整数。

12、根据权利要求9所述的方法，其特征在于，所述负载信息还包括多个执行核心的使用率；

所述方法还包括：

所述转码设备根据所述多个执行核心的使用率，确定平均CPU使用率；

所述转码设备根据所述当前负载值，生成所述状态请求的响应消息，包括：

所述转码设备根据所述当前负载值和所述确定平均CPU使用率，生成所述状态请求的响应消息。

13、根据权利要求12所述的方法，其特征在于，所述转码设备根据所述多个执行核心的使用率，确定平均CPU使用率，包括：

所述转码设备将所述多个执行核心按照每个执行核心的使用率从大到小

进行排名；

所述转码设备将排名前N位的执行核心的使用率的平均值确定为所述平均CPU使用率，N为大于或等于1的整数。

14、根据权利要求9所述的方法，其特征在于，所述状态请求包括所述待分配任务的目标需求信息，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率和所述待分配任务的目标帧率；

所述方法还包括：

所述转码设备根据所述待分配任务的目标需求信息，确定所述待分配任务的负载值；

所述转码设备根据所述当前负载值和所述待分配任务的负载值，确定自身的预测负载值；

所述转码设备根据所述当前负载值，生成所述状态请求的响应消息，包括：

所述转码设备根据所述当前负载值和所述预测负载值，生成所述状态请求的响应消息。

15、根据权利要求9至13中任一项所述的方法，其特征在于，所述状态请求包括所述待分配任务的目标需求信息和所述待分配任务的标识，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率、所述待分配任务的目标帧率；

转码设备在接收到调度设备发送的状态请求之后，在确定所述当前负载值之前，所述方法还包括：

所述转码设备确定所述转码设备承载的转码任务中不存在与所述待分配任务的目标分辨率、所述待分配任务的目标帧率和所述待分配任务的标识均一致的转码任务。

16、一种调度设备，其特征在于，所述调度设备包括：

接收单元，用于接收到客户端设备发送的任务分配请求后，获取多个转码设备的负载信息，每个转码设备的负载信息包括转码设备承载的转码任务的目标

标需求信息,所述转码任务的目标需求信息包括所述转码任务的目标分辨率和所述转码任务的目标帧率;

处理单元,用于针对第一转码设备,根据所述第一转码设备中的至少一个转码任务的目标需求信息以及所述第一转码设备的设备性能,确定所述第一转码设备的当前负载值,所述第一转码设备为所述多个转码设备中的任一设备;以及根据所述多个转码设备的当前负载值,从所述多个转码设备中确定出当前负载值小于第一阈值的目标转码设备;

发送单元,用于向所述客户端设备返回所述任务分配请求的响应消息,所述响应消息用于指示所述客户端设备向所述目标转码设备发送转码任务。

17、根据权利要求16所述的调度设备,其特征在于,所述处理单元具体用于:

根据所述第一转码设备的设备性能,确定所述第一转码设备承载的各转码任务的各目标需求信息对应的预设最大任务数量;

根据所述第一转码设备中每个目标需求信息对应的预设最大任务数量,以及所述第一转码设备中每个目标需求信息对应的转码任务的数量,确定所述第一转码设备的当前负载值。

18、根据权利要求17所述的调度设备,其特征在于,所述当前负载值通过以下方式确定:

$$A=f_1\times\frac{1}{X_1}+f_2\times\frac{1}{X_2}+\cdots+f_n\times\frac{1}{X_n}$$

其中, A 为所述第一转码设备的当前负载值; f_i 为所述第一转码设备中第 i 个目标需求信息对应的任务数量, X_i 为所述第一转码设备中第 i 个目标需求信息对应的预设最大任务数量, $i=1, 2, \dots, n$; n 为所述第一转码设备承载的转码任务的目标需求信息的数目, n 为大于或等于 1 的整数。

19、根据权利要求16所述的调度设备,其特征在于,所述每个转码设备的负载信息还包括所述每个转码设备中多个执行核心的使用率;

所述处理单元具体用于:

根据所述多个转码设备的当前负载值和所述多个转码设备的平均中央处理器CPU使用率，从所述多个转码设备中确定出当前负载值小于第一阈值，且平均CPU使用率小于第二阈值的目标转码设备；所述转码设备的平均CPU使用率是根据转码设备中的多个执行核心的使用率确定的。

20、根据权利要求19所述的调度设备，其特征在于，所述处理单元具体用于：

将所述第一转码设备中的多个执行核心按照每个执行核心的使用率从大到小进行排名；

将排名前N位的执行核心的使用率的平均值确定为所述第一转码设备的平均CPU使用率，N为大于或等于1的整数。

21、根据权利要求19所述的调度设备，其特征在于，所述处理单元具体用于：

从所述多个转码设备中确定出当前负载值小于所述第一阈值的至少一个候选转码设备；

根据每个候选转码设备中的多个执行核心的使用率，确定所述每个候选转码设备的平均CPU使用率；

从所述至少一个候选转码设备中确定出平均CPU使用率小于所述第二阈值的目标转码设备。

22、根据权利要求16所述的调度设备，其特征在于，所述任务分配请求包括待分配任务的目标需求信息，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率和所述待分配任务的目标帧率；

所述处理单元具体用于：

根据所述多个转码设备的当前负载值和所述多个转码设备的预测负载值，从所述多个转码设备中确定出当前负载值小于第一阈值，且预测负载值小于第三阈值的目标转码设备；所述转码设备的预设负载值是根据所述转码设备的当前负载值和所述待分配转码任务的负载值确定的，所述待分配任务的负载值是

根据所述待分配转码任务的目标需求信息确定的。

23、根据权利要求16至22中任一项所述的调度设备，其特征在于，所述任务分配请求包括待分配任务的目标需求信息和所述待分配任务的标识，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率、所述待分配任务的目标帧率；所述每个转码设备的负载信息还包括转码设备承载的转码任务的标识；

所述处理单元在确定所述第一转码设备的当前负载值之前，还用于：

确定所述多个转码设备承载的转码任务中不存在与所述待分配任务的目标分辨率、所述待分配任务的目标帧率和所述待分配任务的标识均一致的转码任务。

24、一种转码设备，其特征在于，所述转码设备包括：

接收单元，用于接收到调度设备发送的状态请求；

处理单元，用于根据自身的负载信息以及自身的设备性能，确定当前负载值；所述负载信息包括所述转码设备承载的至少一个转码任务的目标需求信息，所述转码任务的目标需求信息包括所述转码任务的目标分辨率和所述转码任务的目标帧率；以及根据所述当前负载值，生成所述状态请求的响应消息；

发送单元，用于向所述调度设备发送的所述响应消息，以使所述调度设备根据所述响应消息确定出目标转码设备，所述目标转码设备用于执行待分配的转码任务。

25、根据权利要求24所述的转码设备，其特征在于，所述处理单元具体用于：

根据所述设备性能，确定所承载的各转码任务的各目标需求信息对应的预设最大任务数量；

根据每个目标需求信息对应的预设最大任务数量，以及所述每个目标需求信息对应的转码任务的数量，确定所述转码设备的当前负载值。

26、根据权利要求25所述的转码设备，其特征在于，所述当前负载值通过

以下方式确定：

$$A=f_1\times\frac{1}{X_1}+f_2\times\frac{1}{X_2}+\cdots+f_n\times\frac{1}{X_n}$$

其中，A 为所述转码设备的当前负载值； f_i 为所述转码设备中第 i 个目标需求信息对应的任务数量， X_i 为所述转码设备中第 i 个目标需求信息对应的预设最大任务数量， $i=1, 2, \dots, n$ ； n 为所述转码设备承载的转码任务的目标需求信息的数目， n 为大于或等于 1 的整数。

27、根据权利要求24所述的转码设备，其特征在于，所述负载信息还包括多个执行核心的使用率；

所述处理单元还用于：

根据所述多个执行核心的使用率，确定平均CPU使用率；

根据所述当前负载值和所述确定平均CPU使用率，生成所述状态请求的响应消息。

28、根据权利要求27所述的转码设备，其特征在于，所述处理单元具体用于：

将所述多个执行核心按照每个执行核心的使用率从大到小进行排名；

将排名前N位的执行核心的使用率的平均值确定为所述平均CPU使用率，N为大于或等于1的整数。

29、根据权利要求24所述的转码设备，其特征在于，所述状态请求包括所述待分配任务的目标需求信息，所述待分配任务的目标需求信息包括所述待分配任务的目标分辨率和所述待分配任务的目标帧率；

所述处理单元还用于：

根据所述待分配任务的目标需求信息，确定所述待分配任务的负载值；

根据所述当前负载值和所述待分配任务的负载值，确定自身的预测负载值；

根据所述当前负载值和所述预测负载值，生成所述状态请求的响应消息。

30、根据权利要求24至29中任一项所述的转码设备，其特征在于，所述状态请求包括所述待分配任务的目标需求信息和所述待分配任务的标识，所述待

分配任务的目标需求信息包括所述待分配任务的目标分辨率、所述待分配任务的目标帧率；

所述处理单元在确定所述当前负载值之前，所述处理单元还用于：

确定所述转码设备承载的转码任务中不存在与所述待分配任务的目标分辨率、所述待分配任务的目标帧率和所述待分配任务的标识均一致的转码任务。

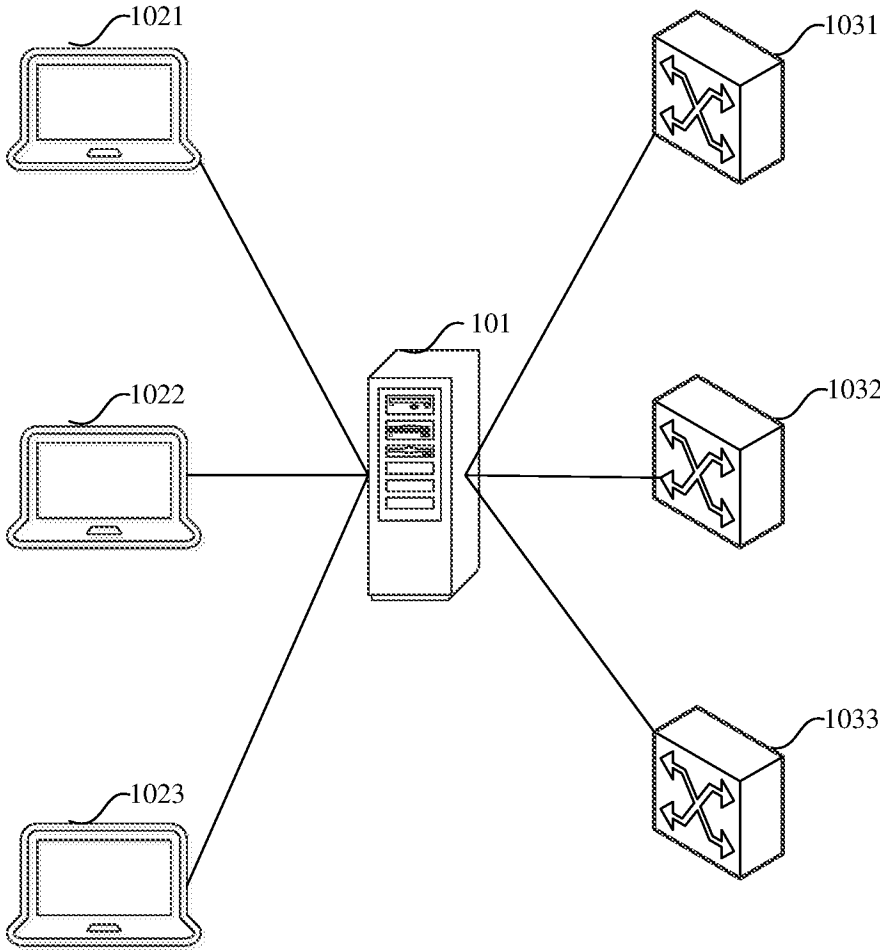


图 1

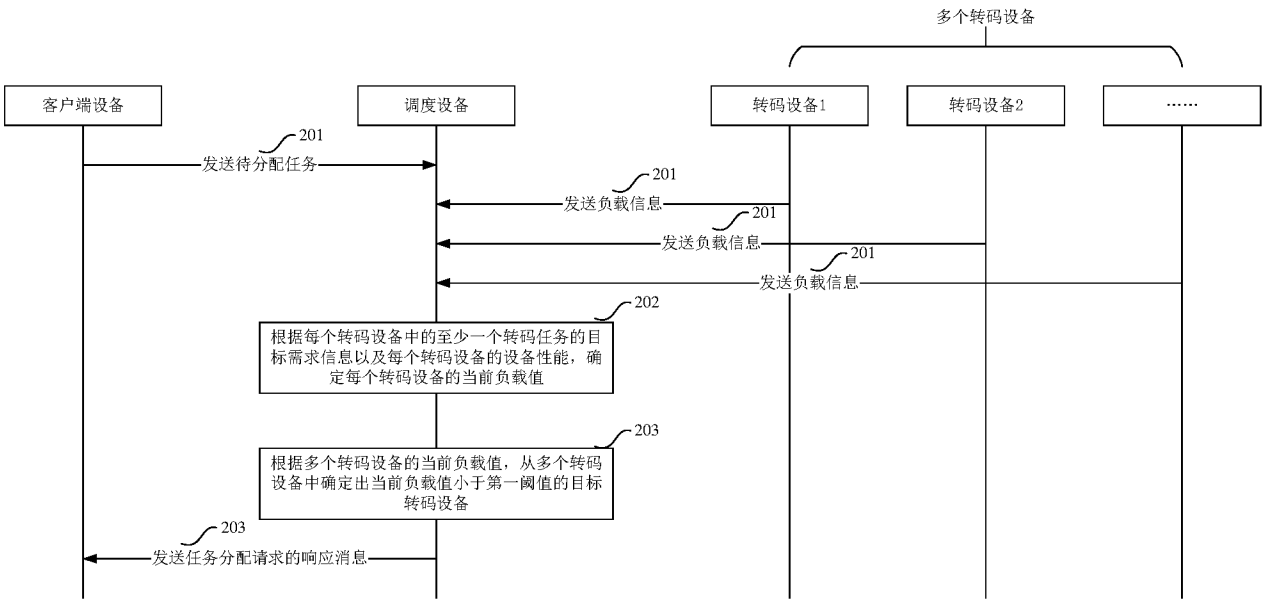


图 2

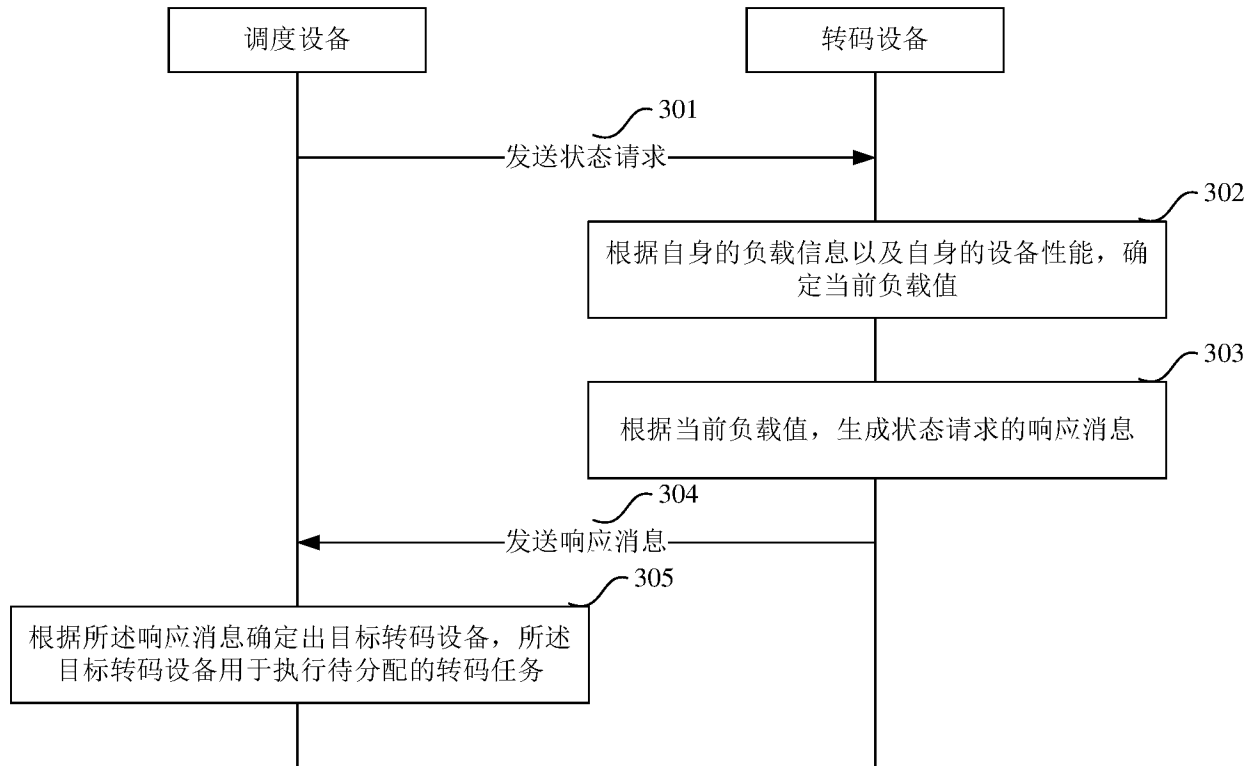


图 3

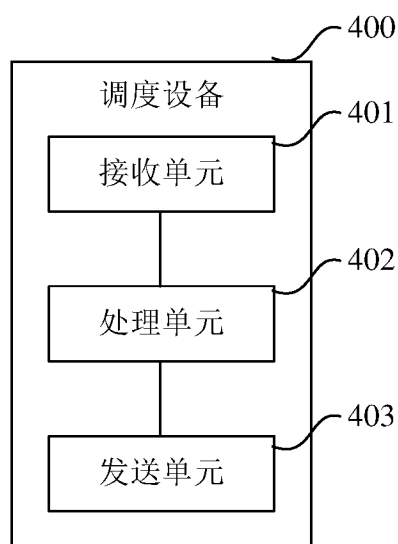


图 4

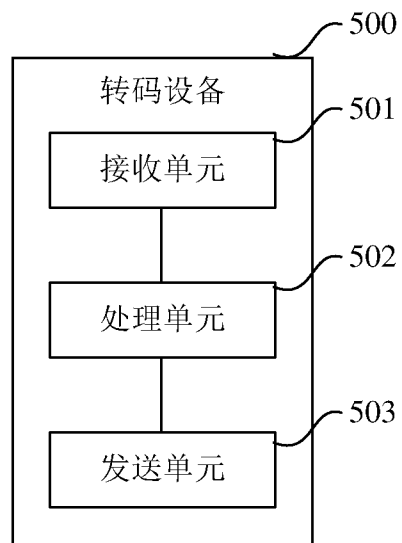


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/107476

A. CLASSIFICATION OF SUBJECT MATTER

H04N 19/40(2014.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNKI, CNABS, CNTXT, VEN, USTXT: 转码, 性能, 负载, 分辨率, 清晰度, 码率, 阈值, 任务, transcode, performance, load, definition, resolution, frame rate, threshold, task, mission

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 105992020 A (LETV CLOUD COMPUTING CO., LTD.) 05 October 2016 (2016-10-05) description, paragraphs 0028-0222, and figures 1-7	1-30
X	CN 107295110 A (WANGSU SCIENCE AND TECHNOLOGY CO., LTD.) 24 October 2017 (2017-10-24) description, paragraphs 0019-0087, and figures 1-5	1-30
X	CN 105162809 A (BEIJING INSTITUTE OF TECHNOLOGY) 16 December 2015 (2015-12-16) description, paragraphs 0026-0054, and figures 1-2	1-30
X	CN 104717517 A (BEIJING QIYI CENTURY SCIENCE & TECHNOLOGY CO., LTD.) 17 June 2015 (2015-06-17) description, paragraphs 0052-0119, and figures 1-3	1-30
X	US 2014267571 A1 (BLUE JEANS NETWORK) 18 September 2014 (2014-09-18) description, paragraphs 0032-0142, and figures 1-9	1-30



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

26 March 2019

Date of mailing of the international search report

04 April 2019

Name and mailing address of the ISA/CN

State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/107476

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	105992020	A	05 October 2016	WO	2017016292	A1	02 February 2017
				US	2017026311	A1	26 January 2017
CN	107295110	A	24 October 2017	None			
CN	105162809	A	16 December 2015	None			
CN	104717517	A	17 June 2015	CN	104717517	B	13 April 2018
US	2014267571	A1	18 September 2014	US	9154735	B2	06 October 2015
				US	2016105641	A1	14 April 2016
				US	9525849	B2	20 December 2016

国际检索报告

国际申请号

PCT/CN2018/107476

A. 主题的分类 H04N 19/40 (2014. 01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类																				
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) H04N 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNKI, CNABS, CNTXT, VEN, USTXT: 转码, 性能, 负载, 分辨率, 清晰度, 码率, 阈值, 任务, transcode, performance, load, definition, resolution, frame rate, threshold, task, mission																				
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 105992020 A (乐视云计算有限公司) 2016年 10月 5日 (2016 - 10 - 05) 说明书第0028-0222段, 图1-7</td> <td>1-30</td> </tr> <tr> <td>X</td> <td>CN 107295110 A (网宿科技股份有限公司) 2017年 10月 24日 (2017 - 10 - 24) 说明书第0019-0087段, 图1-5</td> <td>1-30</td> </tr> <tr> <td>X</td> <td>CN 105162809 A (北京理工大学) 2015年 12月 16日 (2015 - 12 - 16) 说明书第0026-0054段, 图1-2</td> <td>1-30</td> </tr> <tr> <td>X</td> <td>CN 104717517 A (北京奇艺世纪科技有限公司) 2015年 6月 17日 (2015 - 06 - 17) 说明书第0052-0119段, 图1-3</td> <td>1-30</td> </tr> <tr> <td>X</td> <td>US 2014267571 A1 (BLUE JEANS NETWORK) 2014年 9月 18日 (2014 - 09 - 18) 说明书第0032-0142段, 图1-9</td> <td>1-30</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 105992020 A (乐视云计算有限公司) 2016年 10月 5日 (2016 - 10 - 05) 说明书第0028-0222段, 图1-7	1-30	X	CN 107295110 A (网宿科技股份有限公司) 2017年 10月 24日 (2017 - 10 - 24) 说明书第0019-0087段, 图1-5	1-30	X	CN 105162809 A (北京理工大学) 2015年 12月 16日 (2015 - 12 - 16) 说明书第0026-0054段, 图1-2	1-30	X	CN 104717517 A (北京奇艺世纪科技有限公司) 2015年 6月 17日 (2015 - 06 - 17) 说明书第0052-0119段, 图1-3	1-30	X	US 2014267571 A1 (BLUE JEANS NETWORK) 2014年 9月 18日 (2014 - 09 - 18) 说明书第0032-0142段, 图1-9	1-30
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
X	CN 105992020 A (乐视云计算有限公司) 2016年 10月 5日 (2016 - 10 - 05) 说明书第0028-0222段, 图1-7	1-30																		
X	CN 107295110 A (网宿科技股份有限公司) 2017年 10月 24日 (2017 - 10 - 24) 说明书第0019-0087段, 图1-5	1-30																		
X	CN 105162809 A (北京理工大学) 2015年 12月 16日 (2015 - 12 - 16) 说明书第0026-0054段, 图1-2	1-30																		
X	CN 104717517 A (北京奇艺世纪科技有限公司) 2015年 6月 17日 (2015 - 06 - 17) 说明书第0052-0119段, 图1-3	1-30																		
X	US 2014267571 A1 (BLUE JEANS NETWORK) 2014年 9月 18日 (2014 - 09 - 18) 说明书第0032-0142段, 图1-9	1-30																		
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																				
<table border="0"> <tr> <td> * 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 </td> <td> “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件 </td> </tr> </table>			* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																			
国际检索实际完成的日期 2019年 3月 26日		国际检索报告邮寄日期 2019年 4月 4日																		
ISA/CN的名称和邮寄地址 中国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		受权官员 孙一凡 电话号码 86-010-62089961																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/107476

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	105992020	A	2016年 10月 5日	WO	2017016292	A1	2017年 2月 2日
				US	2017026311	A1	2017年 1月 26日
CN	107295110	A	2017年 10月 24日	无			
CN	105162809	A	2015年 12月 16日	无			
CN	104717517	A	2015年 6月 17日	CN	104717517	B	2018年 4月 13日
US	2014267571	A1	2014年 9月 18日	US	9154735	B2	2015年 10月 6日
				US	2016105641	A1	2016年 4月 14日
				US	9525849	B2	2016年 12月 20日